

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
7 June 2001 (07.06.2001)

PCT

(10) International Publication Number  
**WO 01/40493 A2**

(51) International Patent Classification<sup>7</sup>: **C12N 15/85**,  
C07K 14/47, 16/18, C12Q 1/68, G06F 17/30, 17/50,  
A01K 67/027

(FR). **ESSIOUX, Laurent** [FR/FR]; 107, rue de Reuilly,  
F-75012 Paris (FR).

(21) International Application Number: **PCT/IB00/01507**

(74) Common Representative: **GENSET**; Intellectual Prop-  
erty Department, 24, rue Royale, F-75008 Paris (FR).

(22) International Filing Date: 4 October 2000 (04.10.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/168,088 30 November 1999 (30.11.1999) **US**

(71) Applicant (for all designated States except US): **GENSET**  
[FR/FR]; Intellectual Property Department, 24, rue Royale,  
F-75008 Paris (FR).

(81) Designated States (national): **AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.**

(84) Designated States (regional): **ARIPO** patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), **Eurasian** patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), **European** patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), **OAPI** patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **COHEN, Daniel** [FR/FR]; 1, boulevard Richard Wallace, F-92200 Neuilly-sur-Seine (FR). **BLUMENFELD, Marta** [FR/FR]; 5, rue Tagore, F-75013 Paris (FR). **CHU-MAKOV, Ilya** [FR/FR]; 196, rue des Chevrefoilles, F-77000 Vaux-le-Penil (FR). **BOUGUELERET, Lydie** [FR/FR]; 108, avenue Victor Hugo, F-92170 Vanves

Published:

— Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/40493 A2

(54) Title: **SCHIZOPHRENIA ASSOCIATED GENE, PROTEINS AND BIALLELIC MARKERS**

(57) Abstract: The invention concerns the human g35030 gene, polynucleotides, polypeptides biallelic markers, and human chromosome 13q31-q33 biallelic markers. The invention also concerns the association established between schizophrenia and bipolar disorder and the biallelic markers and the g35030 gene and nucleotide sequences. The invention provides means to identify compounds useful in the treatment of schizophrenia, bipolar disorder and related diseases, means to determine the predisposition of individuals to said disease as well as means for the disease diagnosis and prognosis.

## SCHIZOPHRENIA ASSOCIATED GENE, PROTEINS AND BIALLELIC MARKERS

5

### FIELD OF THE INVENTION

The invention concerns the human g35030 gene, and g35030 polynucleotides, polypeptides, as well as biallelic markers localized in the g35030 gene and in the human chromosome 13q31-q33 region. The invention also concerns the association established  
10 between schizophrenia and bipolar disorder and the biallelic markers and the g35030 gene and nucleotide sequences. The invention provides means to identify compounds useful in the treatment of schizophrenia, bipolar disorder and related diseases, means to determine the predisposition of individuals to said disease as well as means for the disease diagnosis and prognosis.

15

### BACKGROUND OF THE INVENTION

Advances in the technological armamentarium available to basic and clinical investigators have enabled increasingly sophisticated studies of brain and nervous system function in health and disease. Numerous hypotheses both neurobiological and  
20 pharmacological have been advanced with respect to the neurochemical and genetic mechanisms involved in central nervous system (CNS) disorders, including psychiatric disorders and neurodegenerative diseases. However, CNS disorders have complex and poorly understood etiologies, as well as symptoms that are overlapping, poorly characterized, and difficult to measure. As a result future treatment regimes and drug development efforts will be  
25 required to be more sophisticated and focused on multigenic causes, and will need new assays to segment disease populations, and provide more accurate diagnostic and prognostic information on patients suffering from CNS disorders.

#### **Genetic Analysis of Complex Traits**

30 Until recently, the identification of genes linked with detectable traits has relied mainly on a statistical approach called linkage analysis. Linkage analysis is based upon establishing a correlation between the transmission of genetic markers and that of a specific trait throughout generations within a family. Linkage analysis involves the study of families with multiple affected individuals and is useful in the detection of inherited-traits, which are caused by a  
35 single gene, or possibly a very small number of genes. But, linkage studies have proven

difficult when applied to complex genetic traits. Most traits of medical relevance do not follow simple Mendelian monogenic inheritance. However, complex diseases often aggregate in families, which suggests that there is a genetic component to be found. Such complex traits are often due to the combined action of multiple genes as well as environmental factors. Such  
5 complex trait, include susceptibilities to heart disease, hypertension, diabetes, cancer and inflammatory diseases. Drug efficacy, response and tolerance/toxicity can also be considered as multifactoral traits involving a genetic component in the same way as complex diseases. Linkage analysis cannot be applied to the study of such traits for which no large informative families are available. Moreover, because of their low penetrance, such complex traits do not  
10 segregate in a clear-cut Mendelian manner as they are passed from one generation to the next. Attempts to map such diseases have been plagued by inconclusive results, demonstrating the need for more sophisticated genetic tools.

Knowledge of genetic variation in the neuronal and endocrine systems is important for understanding why some people are more susceptible to disease or respond differently to  
15 treatments. Ways to identify genetic polymorphism and to analyze how they impact and predict disease susceptibility and response to treatment are needed.

Although the genes involved in the neuronal and endocrine systems represent major drug targets and are of high relevance to pharmaceutical research, we still have scant knowledge concerning the extent and nature of, sequence variation in these genes and their regulatory  
20 elements. In the case where polymorphisms have been identified the relevance of the variation is rarely understood. While polymorphisms hold promise for use as genetic markers in determining which genes contribute to multigenic or quantitative traits, suitable markers and suitable methods for exploiting those markers have not been found and brought to bare on the genes related to disorders of the brain and nervous system.

25 The basis for accomplishment of these goals is to use genetic association analysis to detect markers that predict susceptibility for these traits. Recently, advances in the fields of genetics and molecular biology have allowed identification of forms, or alleles, of human genes that lead to diseases. Most of the genetic variations responsible for human diseases identified so far, belong to the class of single gene disorders. As this name implies, the development of  
30 single gene disorders is determined, or largely influenced, by the alleles of a single gene. The alleles that cause these disorders are, in general, highly deleterious (and highly penetrant) to individuals who carry them. Therefore, these alleles and their associated diseases, with some exceptions, tend to be very rare in the human population. In contrast, most common diseases and non-disease traits, such as a physiological response to a pharmaceutical agent, can be  
35 viewed as the result of many complex factors. These can include environmental exposures

(toxins, allergens, infectious agents, climate, and trauma) as well as multiple genetic factors.

Association studies seek to analyze the distributions of chromosomes that have occurred in populations of unrelated (at least not directly related) individuals. An assumption in this type of study is that genetic alleles that result in susceptibility for a common trait arose by ancient mutational events on chromosomes that have been passed down through many generations in the population. These alleles can become common throughout the population in part because the trait they influence, if deleterious, is only expressed in a fraction of those individuals who carry them. Identification of these "ancestral" chromosomes is made difficult by the fact that genetic markers are likely to have become separated from the trait susceptibility allele through the process of recombination, except in regions of DNA which immediately surround the allele. The identities of genetic markers contained within the fragments of DNA surrounding a susceptibility allele will be the same as those from the ancestral chromosome on which the allele arose. Therefore, individuals from the population who express a complex trait might be expected to carry the same set of genetic markers in the vicinity of a susceptibility allele more often than those who do not express the trait; that is these markers will show an association with the trait.

### Schizophrenia

Schizophrenia is one of the most severe and debilitating of the major psychiatric diseases. It usually starts in late adolescence or early adult life and often becomes chronic and disabling. Men and women are at equal risk of developing this illness; however, most males become ill between 16 and 25 years old, while females develop symptoms between 25 and 30. People with schizophrenia often experience both "positive" symptoms (e.g., delusions, hallucinations, disorganized thinking, and agitation) and "negative" symptoms (e.g., lack of drive or initiative, social withdrawal, apathy, and emotional unresponsiveness).

Schizophrenia affects 1% of the world population. There are an estimated 45 million people with schizophrenia in the world, with more than 33 million of them in the developing countries. This disease places a heavy burden on the patient's family and relatives, both in terms of the direct and indirect costs involved and the social stigma associated with the illness, sometimes over generations. Such stigma often leads to isolation and neglect.

Moreover, schizophrenia accounts for one fourth of all mental health costs and takes up one in three psychiatric hospital beds. Most schizophrenia patients are never able to work. The cost of schizophrenia to society is enormous. In the United States, for example, the direct cost of treatment of schizophrenia has been estimated to be close to 0.5% of the gross national product. Standardized mortality ratios (SMRs) for schizophrenic patients are estimated to be two to four times higher than the general population, and their life expectancy overall is 20 %



shorter than for the general population. The most common cause of death among schizophrenic patients is suicide (in 10 % of patients) which represents a 20 times higher risk than for the general population. Deaths from heart disease and from diseases of the respiratory and digestive system are also increased among schizophrenic patients.

5                   **Bipolar Disorder**

Bipolar disorders are relatively common disorders with severe and potentially disabling effects. In addition to the severe effects on patients' social development, suicide completion rates among bipolar patients are reported to be about 15%.

10                   Bipolar disorders are characterized by phases of excitement and often including depression; the excitement phases, referred to as mania or hypomania, and depression can alternate or occur in various admixtures, and can occur to different degrees of severity and over varying time periods. Because bipolar disorders can exist in different forms and display different symptoms, the classification of bipolar disorder has been the subject of extensive studies resulting in the definition of bipolar disorder subtypes and widening of the overall concept to include patients previously thought to be suffering from different disorders. Bipolar disorders often share certain clinical signs, symptoms, treatments and neurobiological features with psychotic illnesses in general and therefore present a challenge to the psychiatrist to make an accurate diagnosis. Furthermore, because the course of bipolar disorders and various mood and psychotic disorders can differ greatly, it is critical to characterize the illness as early as possible in order to offer means to manage the illness over a long term.

20                   Bipolar disorders appear in about 1.3% of the population and have been reported to constitute about half of the mood disorders seen in a psychiatric clinic. Bipolar disorders have been found to vary with gender depending of the type of disorder; for example, bipolar disorder I is found equally among men and women, while bipolar disorder II is reportedly more common in women. The age of onset of bipolar disorders is typically in the teenage years and diagnosis is typically made in the patient's early twenties. Bipolar disorders also occur among the elderly, generally as a result of a medical or neurological disorder.

25                   The costs of bipolar disorders to society are enormous. The mania associated with the disease impairs performance and causes psychosis, and often results in hospitalization. This disease places a heavy burden on the patient's family and relatives, both in terms of the direct and indirect costs involved and the social stigma associated with the illness, sometimes over generations. Such stigma often leads to isolation and neglect. Furthermore, the earlier the onset, the more severe are the effects of interrupted education and social development.

30                   The DSM-IV classification of bipolar disorder distinguishes among four types of disorders based on the degree and duration of mania or hypomania as well as two types of

disorders which are evident typically with medical conditions or their treatments, or to substance abuse. Mania is recognized by elevated, expansive or irritable mood as well as by distractability, impulsive behavior, increased activity, grandiosity, elation, racing thoughts, and pressured speech. Of the four types of bipolar disorder characterized by the particular degree and duration of mania, DSM-IV includes:

- bipolar disorder I, including patients displaying mania for at least one week;
- bipolar disorder II, including patients displaying hypomania for at least 4 days, characterized by milder symptoms of excitement than mania, who have not previously displayed mania, and have previously suffered from episodes of major depression;
- bipolar disorder not otherwise specified (NOS), including patients otherwise displaying features of bipolar disorder II but not meeting the 4 day duration for the excitement phase, or who display hypomania without an episode of major depression; and
- cyclothymia, including patients who show numerous manic and depressive symptoms that do not meet the criteria for hypomania or major depression, but which are displayed for over two years without a symptom-free interval of more than two months.

The remaining two types of bipolar disorder as classified in DSM-VI are disorders evident or caused by various medical disorder and their treatments, and disorders involving or related to substance abuse. Medical disorders which can cause bipolar disorders typically include endocrine disorders and cerebrovascular injuries, and medical treatments causing bipolar disorder are known to include glucocorticoids and the abuse of stimulants. The disorder associated with the use or abuse of a substance is referred to as "substance induced mood disorder with manic or mixed features".

Diagnosis of bipolar disorder can be very challenging. One particularly troublesome difficulty is that some patients exhibit mixed states, simultaneously manic and dysphoric or depressive, but do not fall into the DSM-IV classification because not all required criteria for mania and major depression are met daily for at least one week. Other difficulties include classification of patients in the DSM-IV groups based on duration of phase since patients often cycle between excited and depressive episodes at different rates. In particular, it is reported that the use of antidepressants may alter the course of the disease for the worse by causing "rapid-cycling". Also making diagnosis more difficult is the fact that bipolar patients, particularly at what is known as Stage III mania, share symptoms of disorganized thinking and behavior with bipolar disorder patients. Furthermore, psychiatrists must distinguish between agitated depression and mixed mania; it is common that patients with major depression (14 days or more) exhibit agitation, resulting in bipolar-like features. A yet further complicating factor is that bipolar patients have an exceptionally high rate of substance, particularly alcohol abuse.

While the prevalence of mania in alcoholic patients is low, it is well known that substance abusers can show excited symptoms. Difficulties therefore result for the diagnosis of bipolar patients with substance abuse.

### Treatment

5           As there are currently no cures for bipolar disorder or schizophrenia, the objective of treatment is to reduce the severity of the symptoms, if possible to the point of remission. Due to the similarities in symptoms, schizophrenia and bipolar disorder are often treated with some of the same medicaments. Both diseases are often treated with antipsychotics and neuroleptics.

10           For schizophrenia, for example, antipsychotic medications are the most common and most valuable treatments. There are four main classes of antipsychotic drugs which are commonly prescribed for schizophrenia. The first, neuroleptics, exemplified by chlorpromazine (Thorazine), has revolutionized the treatment of schizophrenic patients by reducing positive (psychotic) symptoms and preventing their recurrence. Patients receiving chlorpromazine have been able to leave mental hospitals and live in community programs or their own homes. But  
15           these drugs are far from ideal. Some 20% to 30% of patients do not respond to them at all, and others eventually relapse. These drugs were named neuroleptics because they produce serious neurological side effects, including rigidity and tremors in the arms and legs, muscle spasms, abnormal body movements, and akathisia (restless pacing and fidgeting). These side effects are so troublesome that many patients simply refuse to take the drugs. Besides, neuroleptics do not  
20           improve the so-called negative symptoms of schizophrenia and the side effects may even exacerbate these symptoms. Thus, despite the clear beneficial effects of neuroleptics, even some patients who have a good short-term response will ultimately deteriorate in overall functioning.

25           The well known deficiencies in the standard neuroleptics have stimulated a search for new treatments and have led to a new class of drugs termed atypical neuroleptics. The first atypical neuroleptic, Clozapine, is effective for about one third of patients who do not respond to standard neuroleptics. It seems to reduce negative as well as positive symptoms, or at least exacerbates negative symptoms less than standard neuroleptics do. Moreover, it has beneficial effects on overall functioning and may reduce the chance of suicide in schizophrenic patients. It  
30           does not produce the troubling neurological symptoms of the standard neuroleptics, or raise blood levels of the hormone prolactin, excess of which may cause menstrual irregularities and infertility in women, impotence or breast enlargement in men. Many patients who cannot tolerate standard neuroleptics have been able to take clozapine. However, clozapine has serious limitations. It was originally withdrawn from the market because it can cause agranulocytosis, a  
35           potentially lethal inability to produce white blood cells. Agranulocytosis remains a threat that

requires careful monitoring and periodic blood tests. Clozapine can also cause seizures and other disturbing side effects (e.g., drowsiness, lowered blood pressure, drooling, bed-wetting, and weight gain). Thus it is usually taken only by patients who do not respond to other drugs.

5 Researchers have developed a third class of antipsychotic drugs that have the virtues of clozapine without its defects. One of these drugs is risperidone (Risperdal). Early studies suggest that it is as effective as standard neuroleptic drugs for positive symptoms and may be somewhat more effective for negative symptoms. It produces more neurological side effects than clozapine but fewer than standard neuroleptics. However, it raises prolactin levels. Risperidone is now prescribed for a broad range of psychotic patients, and many clinicians seem  
10 to use it before clozapine for patients who do not respond to standard drugs, because they regard it as safer. Another new drug is Olanzapine (Zyprexa) which is at least as effective as standard drugs for positive symptoms and more effective for negative symptoms. It has few neurological side effects at ordinary clinical doses, and it does not significantly raise prolactin levels. Although it does not produce most of clozapine's most troubling side effects, including  
15 agranulocytosis, some patients taking olanzapine may become sedated or dizzy, develop dry mouth, or gain weight. In rare cases, liver function tests become transiently abnormal.

Outcome studies in schizophrenia are usually based on hospital treatment studies and may not be representative of the population of schizophrenia patients. At the extremes of outcome, 20 % of patients seem to recover completely after one episode of psychosis, whereas  
20 14-19% of patients develop a chronic unremitting psychosis and never fully recover. In general, clinical outcome at five years seems to follow the rule of thirds: with about 35 % of patients in the poor outcome category; 36 % in the good outcome category, and the remainder with intermediate outcome. Prognosis in schizophrenia does not seem to worsen after five years.

Whatever the reasons, there is increasing evidence that leaving schizophrenia untreated  
25 for long periods early in course of the illness may negatively affect the outcome. However, the use of drugs is often delayed for patients experiencing a first episode of the illness. The patients may not realize that they are ill, or they may be afraid to seek help; family members sometimes hope the problem will simply disappear or cannot persuade the patient to seek treatment; clinicians may hesitate to prescribe antipsychotic medications when the diagnosis is uncertain  
30 because of potential side effects. Indeed, at the first manifestation of the disease, schizophrenia is difficult to distinguish from bipolar manic-depressive disorders, severe depression, drug-related disorders, and stress-related disorders. Since the optimum treatments differ among these diseases, the long term prognosis of the disorder also differs the beginning of the treatment.

For both schizophrenia and bipolar disorder, all the known molecules used for the  
35 treatment of schizophrenia have side effects and act only against the symptoms of the disease.

There is a strong need for new molecules without associated side effects and directed against targets which are involved in the causal mechanisms of schizophrenia and bipolar disorder. Therefore, tools facilitating the discovery and characterization of these targets are necessary and useful.

5           The aggregation of schizophrenia and bipolar disorder in families, the evidence from  
twin and adoption studies, and the lack of variation in incidence worldwide, indicate that  
schizophrenia and bipolar disorder are primarily genetic conditions, although environmental  
risk factors are also involved at some level as necessary, sufficient, or interactive causes. For  
example, schizophrenia occurs in 1% of the general population. But, if there is one grandparent  
10 with schizophrenia, the risk of getting the illness increases to about 3%; one parent with  
Schizophrenia, to about 10%. When both parents have schizophrenia, the risk rises to  
approximately 40%.

Consequently, there is a strong need to identify genes involved in schizophrenia and bipolar disorder. The knowledge of these genes will allow researchers to understand the etiology of schizophrenia and bipolar disorder and could lead to drugs and medications which are directed against the cause of the diseases, not just against their symptoms.

There is also a great need for new methods for detecting a susceptibility to schizophrenia and bipolar disorder, as well as for preventing or following up the development of the disease. Diagnostic tools could also prove extremely useful. Indeed, early identification of subjects at risk of developing schizophrenia would enable early and/or prophylactic treatment to be administered. Moreover, accurate assessments of the eventual efficacy of a medicament as well as the patient's eventual tolerance to it may enable clinicians to enhance the benefit/risk ratio of schizophrenia and bipolar disorder treatment regimes.

25 SUMMARY OF THE INVENTION

The present invention stems from the identification of novel polymorphisms including biallelic markers located on the human chromosome 13q31-q33 locus, the identification and characterization of novel schizophrenia-related genes located on the human chromosome 13q31-q33 locus, and from the identification of genetic associations between alleles of biallelic markers located on the human chromosome 13q31-q33 locus and disease, as confirmed and characterized in a panel of human subjects. The invention furthermore provides a fine structure map of the region which includes the schizophrenia-associated gene sequences.

The present invention pertains to nucleic acid molecules comprising the genomic sequences of novel human genes encoding g35030 proteins, proteins encoded thereby, as well as antibodies thereto. The g35030 genomic sequences may also comprise regulatory sequence

located upstream (5'-end) and downstream (3'-end) of the transcribed portion of said gene, these regulatory sequences being also part of the invention. The invention also deals with the cDNA sequences encoding g35030 proteins.

5 Oligonucleotide probes or primers hybridizing specifically with a g35030 genomic or cDNA sequence are also part of the present invention, as well as DNA amplification and detection methods using said primers and probes.

A further object of the invention consists of recombinant vectors comprising any of the nucleic acid sequences described above, and in particular of recombinant vectors comprising a g35030 regulatory sequence or a sequence encoding a g35030 protein, as well as of cell hosts  
10 and transgenic non human animals comprising said nucleic acid sequences or recombinant vectors.

The invention also concerns to biallelic markers of the g35030 gene and the use thereof. Included are probes and primers for use in genotyping biallelic markers of the invention.

An embodiment of the invention encompasses any polynucleotide of the invention  
15 attached to a solid support polynucleotide may comprise a sequence disclosed in the present specification; optionally, said polynucleotide may comprise, consist of, or consist essentially of any polynucleotide described in the present specification; optionally, said determining may be performed in a hybridization assay, sequencing assay, microsequencing assay, or an enzyme-based mismatch detection assay; optionally, said polynucleotide may be attached to a solid  
20 support, array, or addressable array; optionally, said polynucleotide may be labeled.

Finally, the invention is directed to drug screening assays and methods for the screening of substances for the treatment of schizophrenia, bipolar disorder or a related CNS disorder. One object of the invention deals with animal models of schizophrenia, including mouse, primate, non-human primate bipolar disorder or related CNS disorder. The invention is also  
25 directed to methods for the screening of substances or molecules that modulate the expression of g35030, as well as with methods for the screening of substances or molecules that interact with a g35030 polypeptide, or that modulate the activity of a g35030 polypeptide.

As noted above, certain aspects of the present invention stem from the identification of genetic associations between schizophrenia and bipolar disorder and alleles of biallelic markers  
30 located on the human chromosome 13q31-q33 region, and more particularly on a subregion thereof referred to herein as Region D. The invention provides appropriate tools for establishing further genetic associations between alleles of biallelic markers on the 13q31-13q33 locus and either side effects or benefit resulting from the administration of agents acting on schizophrenia or bipolar disorder, or schizophrenia or bipolar disorder symptoms, including  
35 agents like chlorpromazine, clozapine, risperidone, olanzapine, sertindole, quetiapine and

ziprasidone.

The invention provides appropriate tools for establishing further genetic associations between alleles of biallelic markers on the 13q31-q33 locus and a trait. Methods and products are provided for the molecular detection of a genetic susceptibility in humans to schizophrenia and bipolar disorder. They can be used for diagnosis, staging, prognosis and monitoring of this disease, which processes can be further included within treatment approaches. The invention also provides for the efficient design and evaluation of suitable therapeutic solutions including individualized strategies for optimizing drug usage, and screening of potential new medicament candidates.

Additional embodiments are set forth in the Detailed Description of the Invention and in the Examples.

#### **BRIEF DESCRIPTION OF THE FIGURES**

Figure 1 is a diagram showing the exon structure of the g35030 gene.

Figure 2 is a table demonstrating the statistical significance of allelic frequencies of selected chromosome 13q31-q33 biallelic markers of the invention in sporadic and familial French Canadian schizophrenia cases and controls.

Figure 3 is a table demonstrating the results of a haplotype association analysis between total French Canadian schizophrenia cases and haplotypes which consist of chromosome 13q31-q33 biallelic markers of the invention.

Figure 4 is a table showing the involvement of selected biallelic markers of the invention in statistically significant haplotypes.

Figure 5 is a table demonstrating the results of a haplotype association analysis between French Canadian schizophrenia cases and haplotypes which consist of chromosome 13q31-q33 biallelic markers of the invention.

Figure 6 is a table demonstrating the results of a haplotype association analysis between French Canadian schizophrenia cases and haplotypes which consist of chromosome 13q31-q33 biallelic markers of the invention.

Figures 7A and 7B show the results of a haplotype association analysis (Omnibus LR test value distribution) between schizophrenia cases and haplotypes comprising Region D biallelic markers of the invention.

Figures 8A and 8B show the results of a haplotype association analysis (HaplotMaxM test value distribution) between schizophrenia cases and haplotypes comprising Region D biallelic markers of the invention.

Figures 9A and 9B show the results of a haplotype association analysis (Omnibus LR

test value distribution) between bipolar disorder cases and haplotypes comprising Region D biallelic markers of the invention.

Figures 10A and 10B show the results of a haplotype association analysis (HaploMaxM test value distribution) between bipolar disorder cases and haplotypes comprising Region D biallelic markers of the invention.

Figures 11A and 11B show the results of a haplotype association analysis (HaploMaxS test value distribution) between bipolar disorder cases and haplotypes comprising Region D biallelic markers of the invention.

Figure 12 is a block diagram of an exemplary computer system.

Figure 13 is a flow diagram illustrating one embodiment of a process 200 for comparing a new nucleotide or protein sequence with a database of sequences in order to determine the homology levels between the new sequence and the sequences in the database.

Figure 14 is a flow diagram illustrating one embodiment of a process 250 in a computer for determining whether two sequences are homologous.

Figure 15 is a flow diagram illustrating one embodiment of an identifier process 300 for detecting the presence of a feature in a sequence.

#### **BRIEF DESCRIPTION OF THE SEQUENCES PROVIDED IN THE SEQUENCE LISTING**

SEQ ID No. 1 contains the approximately 319kb of genomic nucleotide sequence comprising g35030 nucleic acid sequences and the biallelic markers A1 to A71 located on the human chromosome 13q31-q33 locus.

SEQ ID Nos. 2 to 17 contain cDNA sequences of the g35030 gene.

SEQ ID Nos. 18 to 23 contain amino acid sequences of g35030 polypeptides, encoded by cDNAs of the invention.

SEQ ID Nos. 24 to 78 respectively contain the nucleotide sequence of the amplicons which comprise the biallelic markers A72 to A127 located on the human chromosome 13q31-q33 locus.

SEQ ID No. 79 to 132 contain genomic nucleotide sequences comprising exons of the g35030 gene from several different primates.

SEQ ID No 133 contains a primer containing the additional PU 5' sequence described further in Example 2.

SEQ ID No 134 contains a primer containing the additional RP 5' sequence described further in Example 2.



In accordance with the regulations relating to Sequence Listings, the following codes have been used in the Sequence Listing to indicate the locations of biallelic markers within the sequences and to identify each of the alleles present at the polymorphic base. The code "r" in the sequences indicates that one allele of the polymorphic base is a guanine, while the other allele is an adenine. The code "y" in the sequences indicates that one allele of the polymorphic base is a thymine, while the other allele is a cytosine. The code "m" in the sequences indicates that one allele of the polymorphic base is an adenine, while the other allele is an cytosine. The code "k" in the sequences indicates that one allele of the polymorphic base is a guanine, while the other allele is a thymine. The code "s" in the sequences indicates that one allele of the polymorphic base is a guanine, while the other allele is a cytosine. The code "w" in the sequences indicates that one allele of the polymorphic base is an adenine, while the other allele is an thymine.

#### **DETAILED DESCRIPTION OF THE INVENTION**

The identification of genes involved in a particular trait such as a specific central nervous system disorder, like schizophrenia, can be carried out through two main strategies currently used for genetic mapping: linkage analysis and association studies. Linkage analysis requires the study of families with multiple affected individuals and is now useful in the detection of mono- or oligogenic inherited traits. Conversely, association studies examine the frequency of marker alleles in unrelated trait (T+) individuals compared with trait negative (T-) controls, and are generally employed in the detection of polygenic inheritance.

##### **Candidate region on the chromosome 13 (linkage analysis)**

Genetic link or "linkage" is based on an analysis of which of two neighboring sequences on a chromosome contains the least recombinations by crossing-over during meiosis. To do this, chromosomal markers, like microsatellite markers, have been localized with precision on the genome. Genetic link analysis calculates the probabilities of recombinations on the target gene with the chromosomal markers used, according to the genealogical tree, the transmission of the disease, and the transmission of the markers. Thus, if a particular allele of a given marker is transmitted with the disease more often than chance would have it (recombination level between 0 and 0.5), it is possible to deduce that the target gene in question is found in the neighborhood of the marker.

Using this technique, it has been possible to localize several genes demonstrating a genetic predisposition of familial cancers. In order to be able to be included in a genetic link study, the families affected by a hereditary form of the disease must satisfy the "informativeness" criteria: several affected subjects (and whose constitutional DNA is available) per generation, and at best having a large number of siblings.

By linkage analysis, observations have been made, according to which a candidate region for schizophrenia is present on chromosome 13q32 locus (Blouin et al., 1998). Linkage analysis has been successfully applied to map simple genetic traits that show clear Mendelian inheritance patterns and which have a high penetrance, but this method suffers from a variety of drawbacks. First, linkage analysis is limited by its reliance on the choice of a genetic model suitable for each studied trait. Furthermore, the resolution attainable using linkage analysis is limited, and complementary studies are required to refine the analysis of the typical 20 Mb regions initially identified through this method. In addition, linkage analysis have proven difficult when applied to complex genetic traits, such as those due to the combined action of multiple genes and/or environmental factors. In such cases, too great an effort and cost are needed to recruit the adequate number of affected families required for applying linkage analysis to these situations. Finally, linkage analysis cannot be applied to the study of traits for which no large informative families are available.

In the present invention alternative means for conducting association studies rather than linkage analysis between markers located on the chromosome 13q31-q33 locus and a trait, preferably schizophrenia or bipolar disorder, are disclosed.

In the present application, additional biallelic markers located on the human chromosome 13q31-q33 locus associated with schizophrenia are disclosed. The identification of these biallelic markers in association with schizophrenia has allowed for the further definition of the chromosomal region suspected of containing a genetic determinant involved in a predisposition to develop schizophrenia and has resulted in the identification of novel gene sequences disclosed herein which are associated with a predisposition to develop schizophrenia. The present invention thus provides an extensive fine structure map of the 13q31-q33 locus, including novel biallelic markers located on the human 13q31-q33 locus, approximately 319kb of genomic nucleotide sequence of a subregion of the human 13q31-q33 locus, and polymorphisms including biallelic markers and nucleotide deletions in said 319kb genomic sequence. The biallelic markers of the human chromosome 13q31-q33 locus and the nucleotide sequences, polymorphisms and gene sequences located in Region D subregion of the human chromosome 13q31-q33 locus are useful as genetic and physical markers for further mapping studies. These identified polymorphisms are used in the design of assays for the reliable detection of genetic susceptibility to schizophrenia and bipolar disorder. They can also be used in the design of drug screening protocols to provide an accurate and efficient evaluation of the therapeutic and side-effect potential of new or already existing medicament or treatment regime.

Using the biallelic markers, the inventors have identified the g35030 candidate schizophrenia gene, further described herein. The sequences comprising the the schizophrenia-

associated genes are useful, for example, for the isolation of other genes in putative gene families, the identification of homologs from other species, for the development of medicaments for treatment of disease, and as probes and primers for diagnostic or screening assays as described herein.

5

### Definitions

As used interchangeably herein, the term "oligonucleotides", and "polynucleotides" include RNA, DNA, or RNA/DNA hybrid sequences of more than one nucleotide in either single chain or duplex form. The term "nucleotide" as used herein as an adjective to describe molecules comprising RNA, DNA, or RNA/DNA hybrid sequences of any length in single-stranded or duplex form. The term "nucleotide" is also used herein as a noun to refer to individual nucleotides or varieties of nucleotides, meaning a molecule, or individual unit in a larger nucleic acid molecule, comprising a purine or pyrimidine, a ribose or deoxyribose sugar moiety, and a phosphate group, or phosphodiester linkage in the case of nucleotides within an oligonucleotide or polynucleotide. Although the term "nucleotide" is also used herein to encompass "modified nucleotides" which comprise at least one modifications (a) an alternative linking group, (b) an analogous form of purine, (c) an analogous form of pyrimidine, or (d) an analogous sugar, for examples of analogous linking groups, purine, pyrimidines, and sugars see for example PCT publication No. WO 95/04064. However, the polynucleotides of the invention are preferably comprised of greater than 50% conventional deoxyribose nucleotides, and most preferably greater than 90% conventional deoxyribose nucleotides. The polynucleotide sequences of the invention may be prepared by any known method, including synthetic, recombinant, *ex vivo* generation, or a combination thereof, as well as utilizing any purification methods known in the art.

As used herein, the term "purified" does not require absolute purity; rather, it is intended as a relative definition. Purification of starting material or natural material is at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated. As an example, purification from 0.1 % concentration to 10 % concentration is two orders of magnitude.

To illustrate, individual cDNA clones isolated from a cDNA library have been conventionally purified to electrophoretic homogeneity. The sequences obtained from these clones could not be obtained directly either from the library or from total human DNA. The cDNA clones are not naturally occurring as such, but rather are obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The conversion of mRNA into a cDNA library involves the creation of a synthetic substance (cDNA) and pure individual

35

cDNA clones can be isolated from the synthetic library by clonal selection. Thus, creating a cDNA library from messenger RNA and subsequently isolating individual clones from that library results in an approximately  $10^4$ - $10^6$  fold purification of the native message.

The term "purified" is further used herein to describe a polypeptide or polynucleotide of the invention which has been separated from other compounds including, but not limited to, polypeptides or polynucleotides, carbohydrates, lipids, etc. The term "purified" may be used to specify the separation of monomeric polypeptides of the invention from oligomeric forms such as homo- or hetero- dimers, trimers, etc. The term "purified" may also be used to specify the separation of covalently closed polynucleotides from linear polynucleotides. A polynucleotide is substantially pure when at least about 50%, preferably 60 to 75% of a sample exhibits a single polynucleotide sequence and conformation (linear versus covalently close). A substantially pure polypeptide or polynucleotide typically comprises about 50%, preferably 60 to 90% weight/weight of a polypeptide or polynucleotide sample, respectively, more usually about 95%, and preferably is over about 99% pure. Polypeptide and polynucleotide purity, or homogeneity, is indicated by a number of means well known in the art, such as agarose or polyacrylamide gel electrophoresis of a sample, followed by visualizing a single band upon staining the gel. For certain purposes higher resolution can be provided by using HPLC or other means well known in the art. As an alternative embodiment, purification of the polypeptides and polynucleotides of the present invention may be expressed as "at least" a percent purity relative to heterologous polypeptides and polynucleotides (DNA, RNA or both). As a preferred embodiment, the polypeptides and polynucleotides of the present invention are at least; 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 96%, 96%, 98%, 99%, or 100% pure relative to heterologous polypeptides and polynucleotides, respectively. As a further preferred embodiment the polypeptides and polynucleotides have a purity ranging from any number, to the thousandth position, between 90% and 100% (e.g., a polypeptide or polynucleotide at least 99.995% pure) relative to either heterologous polypeptides or polynucleotides, respectively, or as a weight/weight ratio relative to all compounds and molecules other than those existing in the carrier. Each number representing a percent purity, to the thousandth position, may be claimed as individual species of purity.

The term "isolated" requires that the material be removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or DNA or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotide could be part of a vector and/or such polynucleotide or polypeptide could be part of a composition, and still be isolated in that the

vector or composition is not part of its natural environment.

The term "primer" denotes a specific oligonucleotide sequence which is complementary to a target nucleotide sequence and used to hybridize to the target nucleotide sequence. A primer serves as an initiation point for nucleotide polymerization catalyzed by either DNA  
5 polymerase, RNA polymerase or reverse transcriptase.

The term "probe" denotes a defined nucleic acid segment (or nucleotide analog segment, e.g., polynucleotide as defined herein) which can be used to identify a specific polynucleotide sequence present in samples, said nucleic acid segment comprising a nucleotide sequence complementary of the specific polynucleotide sequence to be identified.

10 The terms "trait" and "phenotype" are used interchangeably herein and refer to any clinically distinguishable, detectable or otherwise measurable property of an organism such as symptoms of, or susceptibility to a disease for example. Typically the terms "trait" or "phenotype" are used herein to refer to symptoms of, or susceptibility to schizophrenia or bipolar disorder; or to refer to an individual's response to an agent acting on schizophrenia or  
15 bipolar disorder; or to refer to symptoms of, or susceptibility to side effects to an agent acting on schizophrenia or bipolar disorder.

The term "allele" is used herein to refer to variants of a nucleotide sequence. A biallelic polymorphism has two forms. Typically the first identified allele is designated as the original allele whereas other alleles are designated as alternative alleles. Diploid organisms  
20 may be homozygous or heterozygous for an allelic form.

The term "heterozygosity rate" is used herein to refer to the incidence of individuals in a population, which are heterozygous at a particular allele. In a biallelic system the heterozygosity rate is on average equal to  $2P_a(1-P_a)$ , where  $P_a$  is the frequency of the least common allele. In order to be useful in genetic studies a genetic marker should have an  
25 adequate level of heterozygosity to allow a reasonable probability that a randomly selected person will be heterozygous.

The term "genotype" as used herein refers the identity of the alleles present in an individual or a sample. In the context of the present invention a genotype preferably refers to the description of the biallelic marker alleles present in an individual or a sample. The term  
30 "genotyping" a sample or an individual for a biallelic marker involves determining the specific allele or the specific nucleotide(s) carried by an individual at a biallelic marker.

The term "mutation" as used herein refers to a difference in DNA sequence between or among different genomes or individuals which has a frequency below 1%.

The term "haplotype" refers to a combination of alleles present in an individual or a  
35 sample on a single chromosome. In the context of the present invention a haplotype preferably

refers to a combination of biallelic marker alleles found in a given individual and which may be associated with a phenotype.

The term "polymorphism" as used herein refers to the occurrence of two or more alternative genomic sequences or alleles between or among different genomes or individuals.

5 "Polymorphic" refers to the condition in which two or more variants of a specific genomic sequence can be found in a population. A "polymorphic site" is the locus at which the variation occurs. A polymorphism may comprise a substitution, deletion or insertion of one or more nucleotides. A single nucleotide polymorphism is a single base pair change. Typically a single nucleotide polymorphism is the replacement of one nucleotide by another nucleotide at the  
10 polymorphic site. Deletion of a single nucleotide or insertion of a single nucleotide, also give rise to single nucleotide polymorphisms. In the context of the present invention "single nucleotide polymorphism" preferably refers to a single nucleotide substitution. Typically, between different genomes or between different individuals, the polymorphic site may be occupied by two different nucleotides.

15 The terms "biallelic polymorphism" and "biallelic marker" are used interchangeably herein to refer to a polymorphism having two alleles at a fairly high frequency in the population, preferably a single nucleotide polymorphism. A "biallelic marker allele" refers to the nucleotide variants present at a biallelic marker site. Typically the frequency of the less common allele of the biallelic markers of the present invention has been validated to be greater  
20 than 1%, preferably the frequency is greater than 10%, more preferably the frequency is at least 20% (i.e. heterozygosity rate of at least 0.32), even more preferably the frequency is at least 30% (i.e. heterozygosity rate of at least 0.42). A biallelic marker wherein the frequency of the less common allele is 30% or more is termed a "high quality biallelic marker." All of the genotyping, haplotyping, association, and interaction study methods of the invention may  
25 optionally be performed solely with high quality biallelic markers.

The location of nucleotides in a polynucleotide with respect to the center of the polynucleotide are described herein in the following manner. When a polynucleotide has an odd number of nucleotides, the nucleotide at an equal distance from the 3' and 5' ends of the polynucleotide is considered to be "at the center" of the polynucleotide, and any nucleotide  
30 immediately adjacent to the nucleotide at the center, or the nucleotide at the center itself is considered to be "within 1 nucleotide of the center." With an odd number of nucleotides in a polynucleotide any of the five nucleotides positions in the middle of the polynucleotide would be considered to be within 2 nucleotides of the center, and so on. When a polynucleotide has an even number of nucleotides, there would be a bond and not a nucleotide at the center of the  
35 polynucleotide. Thus, either of the two central nucleotides would be considered to be "within 1

nucleotide of the center" and any of the four nucleotides in the middle of the polynucleotide would be considered to be "within 2 nucleotides of the center", and so on. For polymorphisms which involve the substitution, insertion or deletion of 1 or more nucleotides, the polymorphism, allele or biallelic marker is "at the center" of a polynucleotide if the difference between the distance from the substituted, inserted, or deleted polynucleotides of the polymorphism and the 3' end of the polynucleotide, and the distance from the substituted, inserted, or deleted polynucleotides of the polymorphism and the 5' end of the polynucleotide is zero or one nucleotide. If this difference is 0 to 3, then the polymorphism is considered to be "within 1 nucleotide of the center." If the difference is 0 to 5, the polymorphism is considered to be "within 2 nucleotides of the center." If the difference is 0 to 7, the polymorphism is considered to be "within 3 nucleotides of the center," and so on. For polymorphisms which involve the substitution, insertion or deletion of 1 or more nucleotides, the polymorphism, allele or biallelic marker is "at the center" of a polynucleotide if the difference between the distance from the substituted, inserted, or deleted polynucleotides of the polymorphism and the 3' end of the polynucleotide, and the distance from the substituted, inserted, or deleted polynucleotides of the polymorphism and the 5' end of the polynucleotide is zero or one nucleotide. If this difference is 0 to 3, then the polymorphism is considered to be "within 1 nucleotide of the center." If the difference is 0 to 5, the polymorphism is considered to be "within 2 nucleotides of the center." If the difference is 0 to 7, the polymorphism is considered to be "within 3 nucleotides of the center," and so on.

The term "upstream" is used herein to refer to a location which, is toward the 5' end of the polynucleotide from a specific reference point.

The terms "base paired" and "Watson & Crick base paired" are used interchangeably herein to refer to nucleotides which can be hydrogen bonded to one another by virtue of their sequence identities in a manner like that found in double-helical DNA with thymine or uracil residues linked to adenine residues by two hydrogen bonds and cytosine and guanine residues linked by three hydrogen bonds (See Stryer, L., *Biochemistry*, 4th edition, 1995).

The terms "complementary" or "complement thereof" are used herein to refer to the sequences of polynucleotides which is capable of forming Watson & Crick base pairing with another specified polynucleotide throughout the entirety of the complementary region. This term is applied to pairs of polynucleotides based solely upon their sequences and not any particular set of conditions under which the two polynucleotides would actually bind.

The terms "g35030 gene", when used herein, encompasses genomic, mRNA and cDNA sequences encoding a g35030 protein, including the untranslated regulatory regions of the genomic DNA.

As used herein the term "13q31-q33-related biallelic marker" relates to a biallelic marker residing in the human chromosome 13q31-q33 region. The term 13q31-q33-related biallelic marker encompasses all of the biallelic markers disclosed in Table 5 and any biallelic markers in linkage disequilibrium therewith, and any biallelic markers in linkage disequilibrium therewith. The preferred chromosome 13q31-q33-related biallelic marker alleles of the present invention include each one the alleles described in Table 5 individually or in groups consisting of all the possible combinations of the alleles listed.

As used herein the term "g35030-related biallelic marker" relates to a set of biallelic markers in linkage disequilibrium with the g35030 gene or a g35030 nucleotide sequence. The term g35030-related biallelic marker encompasses the biallelic markers A13 to A65 disclosed in Table 5 and any biallelic markers in linkage disequilibrium therewith.

The term "polypeptide" refers to a polymer of amino acids without regard to the length of the polymer; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not specify or exclude post-expression modifications of polypeptides, for example, polypeptides which include the covalent attachment of glycosyl groups, acetyl groups, phosphate groups, lipid groups and the like are expressly encompassed by the term polypeptide. Also included within the definition are polypeptides which contain one or more analogs of an amino acid (including, for example, non-naturally occurring amino acids, amino acids which only occur naturally in an unrelated biological system, modified amino acids from mammalian systems etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

As used herein, the term "non-human animal" refers to any non-human vertebrate, birds and more usually mammals, preferably primates, farm animals such as swine, goats, sheep, donkeys, and horses, rabbits or rodents, more preferably rats or mice. As used herein, the term "animal" is used to refer to any vertebrate, preferable a mammal. Both the terms "animal" and "mammal" expressly embrace human subjects unless preceded with the term "non-human".

As used herein, the term "antibody" refers to a polypeptide or group of polypeptides which are comprised of at least one binding domain, where an antibody binding domain is formed from the folding of variable domains of an antibody molecule to form three-dimensional binding spaces with an internal surface shape and charge distribution complementary to the features of an antigenic determinant of an antigen, which allows an immunological reaction with the antigen. Antibodies include recombinant proteins comprising the binding domains, as well as fragments, including Fab, Fab', F(ab)<sub>2</sub>, and F(ab')<sub>2</sub> fragments.

As used herein, an "antigenic determinant" is the portion of an antigen molecule, in this case a g35030 polypeptide, that determines the specificity of the antigen-antibody reaction. An



"epitope" refers to an antigenic determinant of a polypeptide. An epitope can comprise as few as 3 amino acids in a spatial conformation which is unique to the epitope. Generally an epitope comprises at least 6 such amino acids, and more usually at least 8-10 such amino acids.

Methods for determining the amino acids which make up an epitope include x-ray crystallography, 2-dimensional nuclear magnetic resonance, and epitope mapping e.g. the Pepscan method described by Geysen et al. 1984; PCT Publication No. WO 84/03564; and PCT Publication No. WO 84/03506.

### Variants and Fragments

The invention also relates to variants and fragments of the polynucleotides described herein, particularly of a nucleotide sequence of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132, and particularly of a nucleotide sequence of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 containing one or more biallelic markers and/or other polymorphisms according to the invention.

Variants of polynucleotides, as the term is used herein, are polynucleotides that differ from a reference polynucleotide. A variant of a polynucleotide may be a naturally occurring variant such as a naturally occurring allelic variant, or it may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the polynucleotide may be made by mutagenesis techniques, including those applied to polynucleotides, cells or organisms.

Generally, differences are limited so that the nucleotide sequences of the reference and the variant are closely similar overall and, in many regions, identical.

Variants of polynucleotides according to the invention include, without being limited to, nucleotide sequences which are at least 95% identical to a polynucleotide selected from the group consisting of the nucleotide sequences SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 or to any polynucleotide fragment of at least 8 consecutive nucleotides of a polynucleotide selected from the group consisting of the nucleotide SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132, and preferably at least 99% identical, more particularly at least 99.5% identical, and most preferably at least 99.8% identical to a polynucleotide selected from the group consisting of the nucleotide SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 or to any polynucleotide fragment of at least 30, 35, 40, 50, 70, 80, 100, 250, 500, 1000 or 2000, to the extent that the length is consistent with the particular sequence ID, consecutive nucleotides of a polynucleotide selected from the group consisting of the nucleotide sequences of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132.

Nucleotide changes present in a variant polynucleotide may be silent, which means that they do not alter the amino acids encoded by the polynucleotide. However, nucleotide changes

may also result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence. The substitutions, deletions or additions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions.

A polynucleotide fragment is a polynucleotide having a sequence that is entirely the same as part but not all of a given nucleotide sequence, preferably the nucleotide sequence of a g35030 polynucleotide, and variants thereof, or of a polynucleotide of any of SEQ ID Nos 1 to 17, 24, 25, 28, 74 and 79 to 132, or a polynucleotide comprising one of the biallelic markers A1 to A127, or the complements thereof. Such fragments may be "free-standing", i.e. not part of or fused to other polynucleotides, or they may be comprised within a single larger polynucleotide of which they form a part or region. Indeed, several of these fragments may be present within a single larger polynucleotide. Optionally, such fragments may comprise, consist of, or consist essentially of a contiguous span of at least 8, 10, 12, 15, 18, 20, 25, 30, 35, 40, 50, 70, 80, 100, 250, 500, 1000 or 2000 nucleotides in length of any of SEQ ID Nos 1 to 17, 24, 25, 28, 74 and 79 to 132.

#### Identity Between Nucleic Acids Or Polypeptides

The terms "percentage of sequence identity" and "percentage homology" are used interchangeably herein to refer to comparisons among polynucleotides and polypeptides, and are determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Homology is evaluated using any of the variety of sequence comparison algorithms and programs known in the art. Such algorithms and programs include, but are by no means limited to, TBLASTN, BLASTP, FASTA, TFASTA, and CLUSTALW (Pearson and Lipman, 1988, Proc. Natl. Acad. Sci. USA 85(8):2444-2448; Altschul et al., 1990, J. Mol. Biol. 215(3):403-410; Thompson et al., 1994, Nucleic Acids Res. 22(2):4673-4680; Higgins et al., 1996, Methods Enzymol. 266:383-402; Altschul et al., 1990, J. Mol. Biol. 215(3):403-410; Altschul et al., 1993, Nature Genetics 3:266-272). In a particularly preferred embodiment, protein and nucleic acid sequence homologies are evaluated using the Basic Local Alignment

Search Tool ("BLAST") which is well known in the art (see, e.g., Karlin and Altschul, 1990, Proc. Natl. Acad. Sci. USA 87:2267-2268; Altschul et al., 1990, J. Mol. Biol. 215:403-410; Altschul et al., 1993, Nature Genetics 3:266-272; Altschul et al., 1997, Nuc. Acids Res. 25:3389-3402). In particular, five specific BLAST programs are used to perform the following

5 task:

(1) BLASTP and BLAST3 compare an amino acid query sequence against a protein sequence database;

(2) BLASTN compares a nucleotide query sequence against a nucleotide sequence database;

10 (3) BLASTX compares the six-frame conceptual translation products of a query nucleotide sequence (both strands) against a protein sequence database;

(4) TBLASTN compares a query protein sequence against a nucleotide sequence database translated in all six reading frames (both strands); and

15 (5) TBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are preferably identified (i.e., aligned) by  
20 means of a scoring matrix, many of which are known in the art. Preferably, the scoring matrix used is the BLOSUM62 matrix (Gonnet et al., 1992, Science 256:1443-1445; Henikoff and Henikoff, 1993, Proteins 17:49-61). Less preferably, the PAM or PAM250 matrices may also be used (see, e.g., Schwartz and Dayhoff, eds., 1978, Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure, Washington: National Biomedical  
25 Research Foundation). The BLAST programs evaluate the statistical significance of all high-scoring segment pairs identified, and preferably selects those segments which satisfy a user-specified threshold of significance, such as a user-specified percent homology. Preferably, the statistical significance of a high-scoring segment pair is evaluated using the statistical significance formula of Karlin (see, e.g., Karlin and Altschul, 1990, Proc. Natl. Acad. Sci. USA  
30 87:2267-2268).

The BLAST programs may be used with the default parameters or with modified parameters provided by the user.

#### **Stringent Hybridization Conditions**

By way of example and not limitation, procedures using conditions of high stringency

are as follows: prehybridization of filters containing DNA is carried out for 8 h to overnight at 65°C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 µg/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65°C, the preferred hybridization temperature, in prehybridization mixture containing 100 µg/ml denatured salmon sperm DNA and 5-20 X 10<sup>6</sup> cpm of <sup>32</sup>P-labeled probe. Subsequently, filter washes can be done at 37°C for 1 h in a solution containing 2 x SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA, followed by a wash in 0.1 X SSC at 50°C for 45 min. Following the wash steps, the hybridized probes are detectable by autoradiography. Other conditions of high stringency which may be used are well known in the art and as cited in Sambrook et al., 1989; and Ausubel et al., 1989. These hybridization conditions are suitable for a nucleic acid molecule of about 20 nucleotides in length. There is no need to say that the hybridization conditions described above are to be adapted according to the length of the desired nucleic acid, following techniques well known to the one skilled in the art. The suitable hybridization conditions may for example be adapted according to the teachings disclosed in the book of Hames and Higgins (1985) or in Sambrook et al.(1989).

#### **Genomic Sequences of g35030 polynucleotides**

The present invention concerns genomic DNA sequences of the g35030 gene. As referred to herein, genomic sequences of g35030 are indicated by nucleotide position in the 5' to 3' orientation on SEQ ID No 1.

Particularly preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising, consisting essentially of, or consisting of a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of nucleotide positions 201123 to 247802 of SEQ ID No 1, or the complements thereof. Also encompassed are purified, isolated, or recombinant polynucleotides comprising a nucleotide sequence having at least 70, 75, 80, 85, 90, or 95% nucleotide identity with nucleotide positions 201123 to 247802 of SEQ ID No 1, or a complementary sequence thereto or a fragment thereof. Also encompassed are purified, isolated, or recombinant nucleic acids that hybridize with said span under stringent hybridization conditions as defined above. Nucleic acids of the invention encompass g35030 nucleic acids from any source, including primate, non-human primate, mammalian and human.

The human g35030 gene comprises exons selected from at least 14 different exons or exon forms, referred to herein as exons S, S2, T, U, V, V1, V2, V3, V4, W, W2, X, Y, Z. As indicated in Table 1, exons V and V2 are considered alternative exons, containing overlapping sequence. Likewise, exon W2 contains overlapping sequence with exons W and X.

The nucleotide positions of g35030 exons in SEQ ID No. 1 are detailed below in Table 1. The exon structure of the g35030 gene is further shown in Figure 1.

Table 1

Exon	Position in SEQID1		Intron Position in SEQID1	
S	201123	201234	201235	201122
S2	201123	201560	201561	214675
T	214676	214793	214794	215701
U	215702	215746	215747	216835
V	216836	216994	216995	217670
V2	216836	217077	217078	217670
V1	217671	217764	217765	227654
V4	227655	227736	227737	238714
V3	238715	238919	238920	240439
W	240440	240673	240674	244352 or 241071
W2	240440	241153	241154	244352
X	241072	241291	241292	244352
Y	244353	244561	244562	246272
Z	246273	247802		

5 Thus, the invention embodies purified, isolated, or recombinant polynucleotides comprising a nucleotide sequence selected from the group consisting of the exons of the g35030 gene, or a sequence complementary thereto. Preferred are purified, isolated, or recombinant polynucleotides comprising at least one exon having the nucleotide position ranges listed in Table 1 selected from the group consisting of the exons S, S2, T, U, V, V1, V2, V3, V4, W, 10 W2, X, Y, Z of the g35030 gene, or a complementary sequence thereto or a fragment or a variant thereof. Also encompassed by the invention are purified, isolated, or recombinant nucleic acids comprising a combination of at least two exons of the g35030 gene selected from the group consisting of exons S, S2, T, U, V, V1, V2, V3, V4, W, W2, X, Y, Z, wherein the polynucleotides are arranged within the nucleic acid in the same relative order as in SEQ ID No. 1. 15

Particularly preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100 or 200 nucleotides of SEQ ID No 1, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ 20 ID No 1: 201123 to 201234, 201123 to 201560, 214676 to 214793, 215702 to 215746, 216836

to 216994, 216836 to 217077, 217671 to 217764, 227655 to 227736, 238715 to 238919, 240440 to 240673, 240440 to 241153, 241072 to 241291, 244353 to 244561 and 246273 to 247802, or the complements thereof.

5 Another object of the invention consists of a purified, isolated, or recombinant nucleic acid that hybridizes with a g35030 nucleotide sequence of nucleotide positions 201123 to 201234, 201123 to 201560, 214676 to 214793, 215702 to 215746, 216836 to 216994, 216836 to 217077, 217671 to 217764, 227655 to 227736, 238715 to 238919, 240440 to 240673, 240440 to 241153, 241072 to 241291, 244353 to 244561 and 246273 to 247802 of SEQ ID No 1, or a complementary sequence thereto or a variant thereof, under stringent hybridization  
10 conditions as defined above.

The invention also encompasses a purified, isolated, or recombinant polynucleotide comprising a nucleotide sequence of g35030 having at least 70, 75, 80, 85, 90, or 95% nucleotide identity with a sequence selected from the group consisting of nucleotide positions 201123 to 201234, 201123 to 201560, 214676 to 214793, 215702 to 215746, 216836 to  
15 216994, 216836 to 217077, 217671 to 217764, 227655 to 227736, 238715 to 238919, 240440 to 240673, 240440 to 241153, 241072 to 241291, 244353 to 244561 and 246273 to 247802 of SEQ ID No. 1 or a complementary sequence thereto or a fragment thereof. The nucleotide differences as regards the nucleotide positions 201123 to 201234, 201123 to 201560, 214676 to 214793, 215702 to 215746, 216836 to 216994, 216836 to 217077, 217671 to 217764, 227655  
20 to 227736, 238715 to 238919, 240440 to 240673, 240440 to 241153, 241072 to 241291, 244353 to 244561 and 246273 to 247802 of SEQ ID No. 1 may generally be distributed throughout the nucleic acid.

The present invention further embodies purified, isolated, or recombinant polynucleotides comprising a nucleotide sequence selected from the group consisting of the  
25 introns of the g35030 gene, or a sequence complementary thereto. The present invention thus encompasses the g35030 gene as well as g35030 genomic sequences consisting of, consisting essentially of, or comprising the sequence of nucleotide positions 201235 to 201122, 201561 to 214675, 214794 to 215701, 215747 to 216835, 216995 to 217670, 217078 to 217670, 217765 to 227654, 227737 to 238714, 238920 to 240439, 240674 to 244352, 240674 to 241071,  
30 241154 to 244352, 241292 to 244352 and 244562 to 246272 of SEQ ID No 1, a sequence complementary thereto, as well as fragments and variants thereof.

These nucleic acids, as well as their fragments and variants, may be used as oligonucleotide primers or probes in order to detect the presence of a copy of a gene comprising  
35 a g35030 nucleic acid sequence in a test sample, or alternatively in order to amplify a target nucleotide sequence within a g35030 nucleic acid sequence or adjoining region.

Additional preferred nucleic acids of the invention include isolated, purified, or recombinant g35030 polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100 or 200 nucleotides of nucleotide positions 201123 to 201234, 201123 to 201560, 214676 to 214793, 215702 to 215746, 216836 to 216994, 216836 to 217077, 217671 to 217764, 227655 to 227736, 238715 to 238919, 240440 to 240673, 240440 to 241153, 241072 to 241291, 244353 to 244561, 246273 to 247802, 201123 to 247802, 199122 to 201122, 247803 to 249803 and 199122 to 249803 of SEQ ID No 1, or the complements thereof, wherein said contiguous span comprises at least one biallelic marker. Optionally, said contiguous span comprises a g35030-related biallelic marker. It should be noted that nucleic acid fragments of any size and sequence may also be comprised by the polynucleotides described in this section. Either the original or the alternative allele may be present at said biallelic marker.

As noted above, g35030 gene sequence are indicated by position on SEQ ID No1. Accordingly, in certain embodiments, a g35030 nucleotide sequence may also comprise isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, 1000 or 2000 nucleotides of nucleotide positions 31 to 319608 of SEQ ID No. 1, or the complements thereof. Optionally, said contiguous span comprises a biallelic marker selected from the group consisting of A13 to A18, A20 to A47, A50 to A53, A56, A58, and A60 to A64. Optionally allele 2 is present at the biallelic marker. It should be noted that nucleic acid fragments of any size and sequence may be comprised by the polynucleotides described in this section.

Further nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 300, 400 or 500 nucleotides, to the extent that the length of said span is consistent with the length of the SEQ ID, of SEQ ID Nos. 24, 25, 28 and 74. The invention also encompasses a purified, isolated, or recombinant polynucleotide comprising a nucleotide sequence having at least 70, 75, 80, 85, 90, or 95% nucleotide identity to said span, as well as a purified, isolated, or recombinant nucleic acid that hybridizes with a said span under the stringent hybridization conditions as defined above. Optionally, said contiguous span comprises a biallelic marker selected from the group consisting of A73 to A74, A77 and A124. Optionally allele 2 is present at the biallelic marker. It should be noted that nucleic acid fragments of any size and sequence may be comprised by the polynucleotides described in this section.

Yet further nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60,

70, 80, 90, 100, 150, 200, 500 or 1000 nucleotides, to the extent that said span is consistent with the nucleotide position range, of SEQ ID No 1, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 201188 to 201234, 216836 to 216915 and 246273 to 247781, or the complements thereof, as well as  
5 polynucleotides having at least 70, 75, 80, 85, 90, or 95% nucleotide identity with said span, and polynucleotides capable of hybridizing with said span. In other embodiments, said contiguous span may specifically exclude one or more of the following nucleotide positions of SEQ ID No. 1: 201188 to 201234, 216836 to 216915, 214676 to 214793 and 215702 to 215746.

10 The present invention also comprises a purified or isolated nucleic acid encoding a g35030 protein having the amino acid sequence of any one of SEQ ID Nos 18 to 23 or a peptide fragment or variant thereof.

It should be noted that nucleic acid fragments of any size and sequence may also be comprised by the polynucleotides described in this section.

15 While this section is entitled "Genomic Sequences of g35030," it should be noted that nucleic acid fragments of any size and sequence may also be comprised by the polynucleotides described in this section, flanking the genomic sequences g35030 on either side or between two or more such genomic sequences.

#### ***g35030 cDNA Sequences***

20 The expression of the g35030 gene has been shown to lead to the production of several mRNA species. Several cDNA sequences corresponding to these mRNA are set forth in SEQ ID Nos 2 to 17.

The invention encompasses a purified, isolated, or recombinant nucleic acid comprising a nucleotide sequence selected from the group consisting of SEQ ID Nos 2 to 17,  
25 complementary sequences thereto, splice variants thereof, as well as allelic variants, and fragments thereof. Moreover, preferred polynucleotides of the invention include purified, isolated, or recombinant g35030 cDNAs consisting of, consisting essentially of, or comprising a nucleotide sequence selected from the group consisting of SEQ ID Nos 2 to 17. Particularly preferred nucleic acids of the invention include isolated, purified, or recombinant  
30 polynucleotides comprising a contiguous span of at least 8, 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 75, 80, 100, 200, 500 or 1000 nucleotides, to the extent that the length of said contiguous span is consistent with the length of the SEQ ID, of a nucleotide sequence selected from the group consisting of SEQ ID Nos 2 to 17, or the complements thereof.

It should be noted that nucleic acid fragments of any size and sequence may also be



comprised by the polynucleotides described in this section.

The invention also pertains to a purified or isolated nucleic acid comprising a polynucleotide having at least 70, 80, 85, 90 or 95% nucleotide identity with a polynucleotide selected from the group consisting of SEQ ID Nos 2 to 17, advantageously 99% nucleotide identity, preferably 99.5% nucleotide identity and most preferably 99.8% nucleotide identity with a polynucleotide selected from the group consisting of SEQ ID Nos 2 to 17, or a sequence complementary thereto or a biologically active fragment thereof.

Another object of the invention relates to purified, isolated or recombinant nucleic acids comprising a polynucleotide that hybridizes, under the stringent hybridization conditions defined herein, with a polynucleotide selected from the group consisting of SEQ ID Nos 2 to 17, or a sequence complementary thereto or a variant thereof or a biologically active fragment thereof.

The g35030 cDNA forms of SEQ ID Nos 2 to 17 are further described in Table 2 below. Shown on the Table 2 are the positions of the 5' UTR, the open reading frame (ORF), the 3' UTR and the polyA signal on the respective SEQ ID No. Also shown are the g35030 exons comprising the cDNA form of a particular SEQ ID No.

Table 2

cDNA	SEQ ID No of cDNA	Position range of 5' UTR in SEQ ID		Position range of ORF in SEQ ID		Position range of 3' UTR in SEQ ID		Position range of polyA signal in SEQ ID	
VV1WYZ	2	1	95	96	167	168	2226	2200	2205
VV1V3WYZ	3	1	95	96	167	168	2431	2405	2410
V2WYZ	4	1	95	96	311	312	2215	2189	2194
VWYZ	5	1	95	96	212	213	2132	2106	2111
V2YZ	6	1	95	96	260	261	1981	1955	1960
VV1YZ	7	1	95	96	167	168	1992	1966	1971
VYZ	8	1	95	96	182	183	1898	1872	1877
VV1Z	9	1	95	96	167	168	1783	1757	1762
V2V1WYZ	10	1	95	96	311	312	2309	2283	2288
V2V1YZ	11	1	95	96	311	312	2075	2049	2054
VWZ	12	1	95	96	212	213	1923	1897	1902
VV1V4WYZ	13	1	95	96	167	168	2308	2282	2287
VV1W2YZ	14	1	95	96	167	168	2706	2680	2685
VW2YZ	15	1	95	96	212	213	2612	2586	2591

STUV	16	1	370	371	436	437	434	-	-
S2TUV	17	1	696	697	762	763	760	-	-

The present invention thus further encompasses variant g35030 polynucleotides having at least one nucleotide substitution, insertion or deletion as described further herein. In preferred embodiments described in Table 3, the invention comprises g35030 polynucleotides comprising a variation as described herein by biallelic markers 8-126-286 (also referred to as 8-130-143) and 8-155-258. The invention thus encompasses purified, isolated, or recombinant polynucleotides and polypeptides encoded thereby, wherein the polynucleotides comprise a contiguous span of at least 8, 12, 15, 18, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 100, 150, or 200 nucleotides of SEQ ID No 2 to 17 or the complement thereof, wherein said contiguous span further comprises nucleotide position 132 of any of SEQ ID Nos 2 to 15 and 17, or nucleotide position 407 of SEQ ID No 16. Optionally, an A is present at said nucleotide position; optionally a G is present at said nucleotide position. In a further embodiment, said contiguous span comprises nucleotide position 328 or SEQ ID 3, or the complement thereof, and optionally a C is present at said position, or optionally a T is present at said position.

Table 3

SEQ ID No	cDNA form	Polymorphism reference	Alleles		Position in SEQ ID No 1	Position in SEQ ID No of cDNA	Position in amino acid seq
2	V1WYZ	8-126-286/ 8-130-143	A	G	216967	132	13
3	VV1V3WYZ					132	13
4	V2WYZ					132	13
5	VWYZ					132	13
6	V2YZ					132	13
7	VV1YZ					132	13
8	VYZ					132	13
9	VV1Z					132	13
10	V2V1WYZ					132	13
11	V2V1YZ					132	13
12	VWZ					132	13
13	VV1V4WYZ					132	13
14	VV1W2YZ					132	13
15	VW2YZ					132	13
16	STUV					407	13
17	S2TUV					733	13
3	VV1V3WYZ	8-155-258	C	T	238789	VV1V3WYZ:328	-

While this section is entitled "g35030 cDNA Sequences," it should be noted that nucleic acid fragments of any size and sequence may also be comprised by the polynucleotides described in this section, flanking the genomic sequences of g35030 on either side or between

two or more such genomic sequences.

### G35030 Coding Regions

The *g35030* open reading frame is contained in the corresponding mRNA of a cDNA sequence selected from the group consisting of SEQ ID Nos 2 to 17. The effective *g35030* coding sequence (CDS) may include several forms as indicated above, in some embodiments encompassing isolated, purified, and recombinant polynucleotides which encode a polypeptide comprising a contiguous span of at least 4 amino acids, preferably 6, more preferably at least 8 or 10 amino acids, yet more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID Nos 18 to 23. The effective *g35030* coding sequence (CDS) may comprise the region between the first nucleotide of the ATG codon and the end nucleotide of the stop codon of SEQ ID Nos 2 to 17 as indicated in Table 2 above.

The above disclosed polynucleotide that contains the coding sequence of the *g35030* gene may be expressed in a desired host cell or a desired host organism when this polynucleotide is placed under the control of suitable expression signals. The expression signals may be either the expression signals contained in the regulatory regions in the *g35030* gene of the invention or in contrast the signals may be exogenous regulatory nucleic sequences. Such a polynucleotide, when placed under the suitable expression signals, may also be inserted in a vector for its expression and/or amplification.

### Regulatory Sequences Of *g35030*

As mentioned, the genomic sequence of the *g35030* gene contains regulatory sequences both in the non-coding 5'-flanking region and in the non-coding 3'-flanking region that border the *g35030* coding region containing the exons of the gene.

Polynucleotides derived from the 5' and 3' regulatory regions are useful in order to detect the presence of at least a copy of a *g35030* nucleotide sequence of SEQ ID No 1 or a fragment thereof in a test sample.

The promoter activity of the 5' regulatory regions contained in *g35030* can be assessed as described below.

In order to identify the relevant biologically active polynucleotide fragments or variants of a *g35030* regulatory region, one of skill in the art will refer to Sambrook et al.(1989), which describes the use of a recombinant vector carrying a marker gene (i.e. beta galactosidase, chloramphenicol acetyl transferase, etc.) the expression of which will be detected when placed under the control of a biologically active polynucleotide fragment or variant of the *g35030* sequence of SEQ ID No 1. Genomic sequences located upstream of the first exon of the *g35030* gene are cloned into a suitable promoter reporter vector, such as the pSEAP-Basic, pSEAP-

Enhancer, p $\beta$ gal-Basic, p $\beta$ gal-Enhancer, or pEGFP-1 Promoter Reporter vectors available from Clontech, or pGL2-basic or pGL3-basic promoterless luciferase reporter gene vector from Promega. Briefly, each of these promoter reporter vectors include multiple cloning sites positioned upstream of a reporter gene encoding a readily assayable protein such as secreted alkaline phosphatase, luciferase,  $\beta$  galactosidase, or green fluorescent protein. The sequences upstream of the g35030 coding region are inserted into the cloning sites upstream of the reporter gene in both orientations and introduced into an appropriate host cell. The level of reporter protein is assayed and compared to the level obtained from a vector which lacks an insert in the cloning site. The presence of an elevated expression level in the vector containing the insert with respect to the control vector indicates the presence of a promoter in the insert. If necessary, the upstream sequences can be cloned into vectors which contain an enhancer for increasing transcription levels from weak promoter sequences. A significant level of expression above that observed with the vector lacking an insert indicates that a promoter sequence is present in the inserted upstream sequence.

Promoter sequence within the upstream genomic DNA may be further defined by constructing nested 5' and/or 3' deletions in the upstream DNA using conventional techniques such as Exonuclease III or appropriate restriction endonuclease digestion. The resulting deletion fragments can be inserted into the promoter reporter vector to determine whether the deletion has reduced or obliterated promoter activity, such as described, for example, by Coles et al.(1998). In this way, the boundaries of the promoters may be defined. If desired, potential individual regulatory sites within the promoter may be identified using site directed mutagenesis or linker scanning to obliterate potential transcription factor binding sites within the promoter individually or in combination. The effects of these mutations on transcription levels may be determined by inserting the mutations into cloning sites in promoter reporter vectors. This type of assay is well-known to those skilled in the art and is described in WO 97/17359, US Patent No. 5,374,544; EP 582 796; US Patent No. 5,698,389; US 5,643,746; US Patent No. 5,502,176; and US Patent 5,266,488.

The strength and the specificity of the promoter of the g35030 gene can be assessed through the expression levels of a detectable polynucleotide operably linked to the g35030 promoter in different types of cells and tissues. The detectable polynucleotide may be either a polynucleotide that specifically hybridizes with a predefined oligonucleotide probe, or a polynucleotide encoding a detectable protein, including a g35030 polypeptide or a fragment or a variant thereof. This type of assay is well-known to those skilled in the art and is described in US Patent No. 5,502,176; and US Patent No. 5,266,488. Some of the methods are discussed in more detail below.

Polynucleotides carrying the regulatory elements located at the 5' end and at the 3' end of the g35030 coding region may be advantageously used to control the transcriptional and translational activity of an heterologous polynucleotide of interest.

Thus, the present invention also concerns a purified or isolated nucleic acid comprising  
5 a polynucleotide which is selected from the group consisting of the 5' and 3' regulatory regions of g35030, or a sequence complementary thereto or a biologically active fragment or variant thereof. In one aspect, "5' regulatory region" may comprise the nucleotide sequence located between positions 199122 to 201122 of SEQ ID No 1. In one aspect, "3' regulatory region" may comprise the nucleotide sequence located between positions 247803 to 249803 of SEQ ID  
10 No 1.

The invention also pertains to a purified or isolated nucleic acid comprising a polynucleotide having at least 95% nucleotide identity with a polynucleotide selected from the group consisting of the 5' and 3' regulatory regions, advantageously 99 % nucleotide identity, preferably 99.5% nucleotide identity and most preferably 99.8% nucleotide identity with a  
15 polynucleotide selected from the group consisting of the 5' and 3' regulatory regions, or a sequence complementary thereto or a variant thereof or a biologically active fragment thereof.

Another object of the invention consists of purified, isolated or recombinant nucleic acids comprising a polynucleotide that hybridizes, under the stringent hybridization conditions defined herein, with a polynucleotide selected from the group consisting of the nucleotide  
20 sequences of the 5'- and 3' regulatory regions of g35030, or a sequence complementary thereto or a variant thereof or a biologically active fragment thereof.

Preferred fragments of the 5' regulatory region have a length of about 1500 or 1000 nucleotides, preferably of about 500 nucleotides, more preferably about 400 nucleotides, even more preferably 300 nucleotides and most preferably about 200 nucleotides.

25 Preferred fragments of the 3' regulatory region are at least 50, 100, 150, 200, 300 or 400 bases in length.

"Biologically active" g35030 regulatory polynucleotide derivatives of SEQ ID No 1 are polynucleotides comprising or alternatively consisting in a fragment of said polynucleotide which is functional as a regulatory region for expressing a recombinant polypeptide or a  
30 recombinant polynucleotide in a recombinant cell host. It could act either as an enhancer or as a repressor.

For the purpose of the invention, a nucleic acid or polynucleotide is "functional" as a regulatory region for expressing a recombinant polypeptide or a recombinant polynucleotide if said regulatory polynucleotide contains nucleotide sequences which contain transcriptional and  
35 translational regulatory information, and such sequences are "operably linked" to nucleotide

sequences which encode the desired polypeptide or the desired polynucleotide.

The regulatory polynucleotides of the invention may be prepared from the nucleotide sequence of SEQ ID No 1 by cleavage using suitable restriction enzymes, as described for example in Sambrook et al.(1989). The regulatory polynucleotides may also be prepared by digestion of SEQ ID No 1 by an exonuclease enzyme, such as Bal31 (Wabiko et al., 1986). These regulatory polynucleotides can also be prepared by nucleic acid chemical synthesis, as described elsewhere in the specification.

The g35030 regulatory polynucleotides according to the invention may be part of a recombinant expression vector that may be used to express a coding sequence in a desired host cell or host organism. The recombinant expression vectors according to the invention are described elsewhere in the specification.

A preferred 5'-regulatory polynucleotide of the invention includes the 5'-untranslated region (5'-UTR) of the g35030 cDNA, or a biologically active fragment or variant thereof.

A preferred 3'-regulatory polynucleotide of the invention includes the 3'-untranslated region (3'-UTR) of the g35030 cDNA, or a biologically active fragment or variant thereof.

A further object of the invention consists of a purified or isolated nucleic acid comprising:

a) a nucleic acid comprising a regulatory nucleotide sequence selected from the group consisting of:

(i) a nucleotide sequence comprising a polynucleotide of the g35030 5' regulatory region or a complementary sequence thereto;

(ii) a nucleotide sequence comprising a polynucleotide having at least 95% of nucleotide identity with the nucleotide sequence of the g35030 5' regulatory region or a complementary sequence thereto;

(iii) a nucleotide sequence comprising a polynucleotide that hybridizes under stringent hybridization conditions with the nucleotide sequence of the g35030 5' regulatory region or a complementary sequence thereto; and

(iv) a biologically active fragment or variant of the polynucleotides in (i), (ii) and (iii);

b) a polynucleotide encoding a desired polypeptide or a nucleic acid of interest, operably linked to the nucleic acid defined in (a) above; and

c) optionally, a nucleic acid comprising a 3'- regulatory polynucleotide, preferably a 3'- regulatory polynucleotide of the g35030 gene.

In a specific embodiment of the nucleic acid defined above, said nucleic acid includes the 5'-untranslated region (5'-UTR) of the g35030 cDNA, or a biologically active fragment or variant thereof.

In a second specific embodiment of the nucleic acid defined above, said nucleic acid includes the 3'-untranslated region (3'-UTR) of the g35030 cDNA, or a biologically active fragment or variant thereof.

5 The regulatory polynucleotide of the 5' regulatory region, or its biologically active fragments or variants, is operably linked at the 5'-end of the polynucleotide encoding the desired polypeptide or polynucleotide.

The regulatory polynucleotide of the 3' regulatory region, or its biologically active fragments or variants, is advantageously operably linked at the 3'-end of the polynucleotide encoding the desired polypeptide or polynucleotide.

10 The desired polypeptide encoded by the above-described nucleic acid may be of various nature or origin, encompassing proteins of prokaryotic or eukaryotic origin. Among the polypeptides expressed under the control of a g35030 regulatory region include bacterial, fungal or viral antigens. Also encompassed are eukaryotic proteins such as intracellular proteins, like "house keeping" proteins, membrane-bound proteins, like receptors, and secreted proteins like  
15 endogenous mediators such as cytokines. The desired polypeptide may be the g35030 protein, especially the protein of the amino acid sequences of SEQ ID Nos 18 to 23, or a fragment or a variant thereof.

The desired nucleic acids encoded by the above-described polynucleotide, usually an RNA molecule, may be complementary to a desired coding polynucleotide, for example to the  
20 g35030 coding sequence, and thus useful as an antisense polynucleotide.

Such a polynucleotide may be included in a recombinant expression vector in order to express the desired polypeptide or the desired nucleic acid in host cell or in a host organism. Suitable recombinant vectors that contain a polynucleotide such as described herein are disclosed elsewhere in the specification.

25

#### **Polynucleotide Constructs**

The terms "polynucleotide construct" and "recombinant polynucleotide" are used interchangeably herein to refer to linear or circular, purified or isolated polynucleotides that have been artificially designed and which comprise at least two nucleotide sequences that are  
30 not found as contiguous nucleotide sequences in their initial natural environment. It should be noted that the present invention embodies recombinant vectors comprising any one of the polynucleotides described in the present invention.

**DNA Constructs that Enables Directing Temporal and Spatial Expression of  
35 g35030 Nucleic Acid Sequences in Recombinant Cell Hosts and in Transgenic Animals**

In order to study the physiological and phenotypic consequences of a lack of synthesis of a protein encoded by a nucleotide sequence comprising a g35030 polynucleotide, both at the cell level and at the multi cellular organism level, the invention also encompasses DNA constructs and recombinant vectors enabling a conditional expression of a specific allele of a nucleotide sequence comprising a g35030 polynucleotide and also of a copy of a sequence comprising a nucleotide sequence of a g35030 polynucleotide, or a fragment thereof, harboring substitutions, deletions, or additions of one or more bases. These base substitutions, deletions or additions may be located either in an exon, an intron or a regulatory sequence, in particular a 5' regulatory sequence of a g35030 polynucleotide. In a preferred embodiment, the nucleotide sequence comprising a g35030 polynucleotide further comprises a biallelic marker of the present invention.

A first preferred DNA construct is based on the tetracycline resistance operon *tet* from *E. coli* transposon Tn10 for controlling the expression of a g35030 polynucleotide, such as described by Gossen et al. (1992, 1995) and Furth et al. (1994). Such a DNA construct contains seven *tet* operator sequences from Tn10 (*tet*op) that are fused to either a minimal promoter or a 5'-regulatory sequence of the g35030 polynucleotide, said minimal promoter or said g35030 polynucleotide regulatory sequence being operably linked to a polynucleotide of interest that codes either for a sense or an antisense oligonucleotide or for a polypeptide, including a g35030 polynucleotide-encoded polypeptide or a peptide fragment thereof. This DNA construct is functional as a conditional expression system for the nucleotide sequence of interest when the same cell also comprises a nucleotide sequence coding for either the wild type (*tTA*) or the mutant (*rTA*) repressor fused to the activating domain of viral protein VP16 of herpes simplex virus, placed under the control of a promoter, such as the HCMVIE1 enhancer/promoter or the MMTV-LTR. Indeed, a preferred DNA construct of the invention comprises both the polynucleotide containing the *tet* operator sequences and the polynucleotide containing a sequence coding for the *tTA* or the *rTA* repressor.

In a specific embodiment, the conditional expression DNA construct contains the sequence encoding the mutant tetracycline repressor *rTA*, the expression of the polynucleotide of interest is silent in the absence of tetracycline and induced in its presence.

#### **DNA Constructs Allowing Homologous Recombination: Replacement Vectors**

A second preferred DNA construct will comprise, from 5'-end to 3'-end: (a) a first nucleotide sequence comprising a g35030 polynucleotide; (b) a nucleotide sequence comprising a positive selection marker, such as the marker for neomycine resistance (*neo*); and (c) a second nucleotide sequence comprising a respective g35030 polynucleotide, and is located on the genome downstream of the first g35030 polynucleotide sequence (a).



In a preferred embodiment, this DNA construct also comprises a negative selection marker located upstream the nucleotide sequence (a) or downstream the nucleotide sequence (c). Preferably, the negative selection marker comprises the thymidine kinase (*tk*) gene (Thomas et al., 1986), the hygromycin beta gene (Te Riele et al., 1990), the *hprt* gene (Van der Lugt et al., 1991; Reid et al., 1990) or the Diphtheria toxin A fragment (*Dt-A*) gene (Nada et al., 1993; Yagi et al. 1990). Preferably, the positive selection marker is located within and exon of a g35030 polynucleotide so as to interrupt the sequence encoding the g35030 protein. These replacement vectors are described, for example, by Thomas et al. (1986; 1987), Mansour et al. (1988) and Koller et al. (1992).

The first and second nucleotide sequences (a) and (c) may be indifferently located within a g35030 polynucleotide regulatory sequence, an intronic sequence, an exon sequence or a sequence containing both regulatory and/or intronic and/or exon sequences. The size of the nucleotide sequence of (a) and (c) ranges from 1 to 50 kb, preferably from 1 to 10 kb, more preferably from 2 to 6 kb and most preferably from 2 to 4 kb.

#### **DNA Constructs Allowing Homologous Recombination: Cre-LoxP System.**

These new DNA constructs make use of the site specific recombination system of the P1 phage. The P1 phage possesses a recombinase called Cre which interacts specifically with a 34 base pairs *loxP* site. The *loxP* site is composed of two palindromic sequences of 13 bp separated by a 8 bp conserved sequence (Hoess et al., 1986). The recombination by the Cre enzyme between two *loxP* sites having an identical orientation leads to the deletion of the DNA fragment.

The Cre-*loxP* system used in combination with a homologous recombination technique has been first described by Gu et al. (1993, 1994). Briefly, a nucleotide sequence of interest to be inserted in a targeted location of the genome harbors at least two *loxP* sites in the same orientation and located at the respective ends of a nucleotide sequence to be excised from the recombinant genome. The excision event requires the presence of the recombinase (Cre) enzyme within the nucleus of the recombinant cell host. The recombinase enzyme may be brought at the desired time either by (a) incubating the recombinant cell hosts in a culture medium containing this enzyme, by injecting the Cre enzyme directly into the desired cell, such as described by Araki et al. (1995), or by lipofection of the enzyme into the cells, such as described by Baubonis et al. (1993); (b) transfecting the cell host with a vector comprising the Cre coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter being optionally inducible, said vector being introduced in the recombinant cell

host, such as described by Gu et al.(1993) and Sauer et al.(1988); (c) introducing in the genome of the cell host a polynucleotide comprising the *Cre* coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter is optionally inducible, and said polynucleotide being inserted in the genome of the cell host either by a random insertion event or an homologous recombination event, such as described by Gu et al.(1994).

In a specific embodiment, the vector containing the sequence to be inserted in a g35030 gene sequence by homologous recombination is constructed in such a way that selectable markers are flanked by *loxP* sites of the same orientation, it is possible, by treatment by the *Cre* enzyme, to eliminate the selectable markers while leaving the g35030 polynucleotide sequences of interest that have been inserted by an homologous recombination event. Again, two selectable markers are needed: a positive selection marker to select for the recombination event and a negative selection marker to select for the homologous recombination event. Vectors and methods using the *Cre-loxP* system are described by Zou et al.(1994).

Thus, in one aspect, a further preferred DNA construct of the invention comprises, from 5'-end to 3'-end: (a) a first nucleotide sequence that is comprised by a g35030 polynucleotide; (b) a nucleotide sequence comprising a polynucleotide encoding a positive selection marker, said nucleotide sequence comprising additionally two sequences defining a site recognized by a recombinase, such as a *loxP* site, the two sites being placed in the same orientation; and (c) a second nucleotide sequence comprising a g35030 polynucleotide, and is located on the genome downstream of the first g35030 polynucleotide sequence (a).

The sequences defining a site recognized by a recombinase, such as a *loxP* site, are preferably located within the nucleotide sequence (b) at suitable locations bordering the nucleotide sequence for which the conditional excision is sought. In one specific embodiment, two *loxP* sites are located at each side of the positive selection marker sequence, in order to allow its excision at a desired time after the occurrence of the homologous recombination event.

In a preferred embodiment of a method using the third DNA construct described above, the excision of the polynucleotide fragment bordered by the two sites recognized by a recombinase, preferably two *loxP* sites, is performed at a desired time, due to the presence within the genome of the recombinant host cell of a sequence encoding the *Cre* enzyme operably linked to a promoter sequence, preferably an inducible promoter, more preferably a tissue-specific promoter sequence and most preferably a promoter sequence which is both inducible and tissue-specific, such as described by Gu et al.(1994).

The presence of the *Cre* enzyme within the genome of the recombinant cell host may result from the breeding of two transgenic animals, the first transgenic animal bearing the g35030 polynucleotide -derived sequence of interest containing the *loxP* sites as described

above and the second transgenic animal bearing the *Cre* coding sequence operably linked to a suitable promoter sequence, such as described by Gu et al.(1994).

5 Spatio-temporal control of the *Cre* enzyme expression may also be achieved with an adenovirus based vector that contains the *Cre* gene thus allowing infection of cells, or *in vivo* infection of organs, for delivery of the *Cre* enzyme, such as described by Anton and Graham (1995) and Kanegae et al.(1995).

10 The DNA constructs described above may be used to introduce a desired nucleotide sequence of the invention, preferably a g35030 polynucleotide, and most preferably an altered copy of a g35030 polynucleotide sequence, within a predetermined location of the targeted genome, leading either to the generation of an altered copy of a targeted gene (knock-out homologous recombination) or to the replacement of a copy of the targeted gene by another copy sufficiently homologous to allow an homologous recombination event to occur (knock-in homologous recombination). In a specific embodiment, the DNA constructs described above may be used to introduce a g35030 polynucleotide.

15 **Nuclear Antisense DNA Constructs**

Other compositions containing a vector of the invention comprise an oligonucleotide fragment of the g35030 polynucleotide sequences of SEQ ID No.1 respectively, as an antisense tool that inhibits the expression of the corresponding gene. Preferred methods using antisense polynucleotide according to the present invention are the procedures described by Sczakiel et al.(1995) or those described in PCT Application No WO 95/24223.

20 Preferably, the antisense tools are chosen among the polynucleotides (15-200 bp long) that are complementary to the 5'end of a g35030 polynucleotide mRNA. In one embodiment, a combination of different antisense polynucleotides complementary to different parts of the desired targeted gene are used.

25 Preferably, the antisense polynucleotides of the invention have a 3' polyadenylation signal that has been replaced with a self-cleaving ribozyme sequence, such that RNA polymerase II transcripts are produced without poly(A) at their 3' ends, these antisense polynucleotides being incapable of export from the nucleus, such as described by Liu et al.(1994). In a preferred embodiment, these g35030 antisense polynucleotides also comprise, 30 within the ribozyme cassette, a histone stem-loop structure to stabilize cleaved transcripts against 3'-5' exonucleolytic degradation, such as the structure described by Eckner et al.(1991).

**Constructs for Activation of g35030 gene**

35 The present invention also encompasses primary, secondary, and immortalized homologously recombinant host cells of vertebrate origin, preferably mammalian origin and particularly human origin, that have been engineered to: a) insert exogenous (heterologous)

polynucleotides into the endogenous chromosomal DNA of a targeted gene, b) delete endogenous chromosomal DNA, and/or c) replace endogenous chromosomal DNA with exogenous polynucleotides. Insertions, deletions, and/or replacements of polynucleotide sequences may be to the coding sequences of the targeted gene and/or to regulatory regions, such as promoter and enhancer sequences, operably associated with the targeted gene.

The present invention further relates to a method of making a homologously recombinant host cell in vitro or in vivo, wherein the expression of a targeted gene not normally expressed in the cell is altered. Preferably the alteration causes expression of the targeted gene under normal growth conditions or under conditions suitable for producing the polypeptide encoded by the targeted gene. The method comprises the steps of: (a) transfecting the cell in vitro or in vivo with a polynucleotide construct, the a polynucleotide construct comprising: (i) a targeting sequence; (ii) a regulatory sequence and/or a coding sequence; and (iii) an unpaired splice donor site, if necessary, thereby producing a transfected cell; and (b) maintaining the transfected cell in vitro or in vivo under conditions appropriate for homologous recombination.

The present invention further relates to a method of altering the expression of a targeted gene in a cell in vitro or in vivo wherein the gene is not normally expressed in the cell, comprising the steps of: (a) transfecting the cell in vitro or in vivo with a polynucleotide construct, the a polynucleotide construct comprising: (i) a targeting sequence; (ii) a regulatory sequence and/or a coding sequence; and (iii) an unpaired splice donor site, if necessary, thereby producing a transfected cell; and (b) maintaining the transfected cell in vitro or in vivo under conditions appropriate for homologous recombination, thereby producing a homologously recombinant cell; and (c) maintaining the homologously recombinant cell in vitro or in vivo under conditions appropriate for expression of the gene.

The present invention further relates to a method of making a polypeptide of the present invention by altering the expression of a targeted endogenous gene in a cell in vitro or in vivo wherein the gene is not normally expressed in the cell, comprising the steps of: a) transfecting the cell in vitro with a a polynucleotide construct, the a polynucleotide construct comprising: (i) a targeting sequence; (ii) a regulatory sequence and/or a coding sequence; and (iii) an unpaired splice donor site, if necessary, thereby producing a transfected cell; (b) maintaining the transfected cell in vitro or in vivo under conditions appropriate for homologous recombination, thereby producing a homologously recombinant cell; and c) maintaining the homologously recombinant cell in vitro or in vivo under conditions appropriate for expression of the gene thereby making the polypeptide.

The present invention further relates to a a polynucleotide construct which alters the expression of a targeted gene in a cell type in which the gene is not normally expressed. This

occurs when the a polynucleotide construct is inserted into the chromosomal DNA of the target cell, wherein the a polynucleotide construct comprises: a) a targeting sequence; b) a regulatory sequence and/or coding sequence; and c) an unpaired splice-donor site, if necessary. Further included are a polynucleotide constructs, as described above, wherein the construct further comprises A a polynucleotide which encodes a polypeptide and is in-frame with the targeted endogenous gene after homologous recombination with chromosomal DNA.

The compositions may be produced, and methods performed, by techniques known in the art, such as those described in U.S. Patent Nos: 6,054,288; 6,048,729; 6,048,724; 6,048,524; 5,994,127; 5,968,502; 5,965,125; 5,869,239; 5,817,789; 5,783,385; 5,733,761; 5,641,670; 5,580,734 ; International Publication Nos: WO96/29411, WO 94/12650; and scientific articles including 1994; Koller et al., Proc. Natl. Acad. Sci. USA 86:8932-8935 (1989) (the disclosures of each of which are incorporated by reference in their entireties).

#### Oligonucleotide Probes And Primers

The polynucleotides of the invention are useful in order to detect the presence of at least a copy of a g35030 nucleotide sequence of SEQ ID Nos 1 to 17 or of the respective g35030 polynucleotide or gene, or a fragment, complement, or variant thereof in a test sample.

Particularly preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, 1000 or 2000 nucleotides, to the extent that said span is consistent with the length of the nucleotide position range, of SEQ ID No 1, wherein said contiguous span comprises at least 1, 2, 3, 4, 5, 7 or 10 of the following nucleotide positions of SEQ ID No 1: 201123 to 201234, 201123 to 201560, 214676 to 214793, 215702 to 215746, 216836 to 216994, 216836 to 217077, 217671 to 217764, 227655 to 227736, 238715 to 238919, 240440 to 240673, 240440 to 241153, 241072 to 241291, 244353 to 244561, 246273 to 247802, 201123 to 247802, 199122 to 201122, 247803 to 249803 and 199122 to 249803, or a complementary sequence thereto or a fragment thereof.

Probes and primers of the invention also include isolated, purified, or recombinant polynucleotides having at least 70, 75, 80, 85, 90, or 95% nucleotide identity with a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, 1000 or 2000 nucleotides of nucleotide positions 201123 to 201234, 201123 to 201560, 214676 to 214793, 215702 to 215746, 216836 to 216994, 216836 to 217077, 217671 to 217764, 227655 to 227736, 238715 to 238919, 240440 to 240673, 240440 to 241153, 241072 to 241291, 244353 to 244561, 246273 to 247802, 201123 to 247802, 199122 to 201122, 247803 to 249803 and 199122 to 249803 of SEQ ID No. 1. Preferred probes and primers of the invention also

include isolated, purified, or recombinant polynucleotides comprising a g35030 nucleotide sequence having at least 70, 75, 80, 85, 90, or 95% nucleotide identity with at least one sequence selected from the group consisting of the following nucleotide positions of SEQ ID No. 1: 201123 to 201234, 201123 to 201560, 214676 to 214793, 215702 to 215746, 216836 to 216994, 216836 to 217077, 217671 to 217764, 227655 to 227736, 238715 to 238919, 240440 to 240673, 240440 to 241153, 241072 to 241291, 244353 to 244561, 246273 to 247802, 201123 to 247802, 199122 to 201122, 247803 to 249803 and 199122 to 249803; or a complementary sequence thereto or a fragment thereof.

The invention also relates to nucleic acid probes characterized in that they hybridize specifically, under the stringent hybridization conditions defined above, with a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, 1000 or 2000 nucleotides of nucleotide positions 201123 to 201234, 201123 to 201560, 214676 to 214793, 215702 to 215746, 216836 to 216994, 216836 to 217077, 217671 to 217764, 227655 to 227736, 238715 to 238919, 240440 to 240673, 240440 to 241153, 241072 to 241291, 244353 to 244561, 246273 to 247802, 201123 to 247802, 199122 to 201122, 247803 to 249803 and 199122 to 249803 of SEQ ID No. 1, or a variant thereof or a sequence complementary thereto.

Probes and primers of the invention further include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, 1000 or 2000 nucleotides, to the extent that said span is consistent with the length of the nucleotide position range, of SEQ ID No 2 to 17 or 79 to 132. Also included are isolated, purified, or recombinant polynucleotides having at least 70, 75, 80, 85, 90, or 95% nucleotide identity with a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, 1000 or 2000 nucleotides of SEQ ID No. 2 to 17 or 79 to 132, or a complementary sequence thereto, as well as isolated, purified, or recombinant polynucleotides comprising a g35030 nucleotide sequence having at least 70, 75, 80, 85, 90, or 95% nucleotide identity with at least one sequence selected from the group consisting of the following nucleotide positions of SEQ ID No. 2 to 17 or 79 to 132; or a complementary sequence thereto or a fragment thereof. The invention also relates to nucleic acid probes characterized in that they hybridize specifically, under the stringent hybridization conditions defined above, with a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, 1000 or 2000 nucleotides of nucleotide SEQ ID No. 2 to 17 or 79 to 132, or a variant thereof or a sequence complementary thereto.

The formation of stable hybrids depends on the melting temperature ( $T_m$ ) of the DNA. The  $T_m$  depends on the length of the primer or probe, the ionic strength of the solution and the G+C content. The higher the G+C content of the primer or probe, the higher is the melting

temperature because G:C pairs are held by three H bonds whereas A:T pairs have only two. The GC content in the probes of the invention usually ranges between 10 and 75 %, preferably between 35 and 60 %, and more preferably between 40 and 55 %.

5 A probe or a primer according to the invention may be between 8 and 2000 nucleotides in length, or is specified to be at least 12, 15, 18, 20, 25, 35, 40, 50, 60, 70, 80, 100, 250, 500 , 1000 nucleotides in length. More particularly, the length of these probes can range from 8, 10, 15, 20, or 30 to 100 nucleotides, preferably from 10 to 50, more preferably from 15 to 30 nucleotides. Shorter probes tend to lack specificity for a target nucleic acid sequence and generally require cooler temperatures to form sufficiently stable hybrid complexes with the  
10 template. Longer probes are expensive to produce and can sometimes self-hybridize to form hairpin structures. The appropriate length for primers and probes under a particular set of assay conditions may be empirically determined by one of skill in the art.

The primers and probes can be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences and direct chemical synthesis by a  
15 method such as the phosphodiester method of Narang et al.(1979), the phosphodiester method of Brown et al.(1979), the diethylphosphoramidite method of Beaucage et al.(1981) and the solid support method described in EP 0 707 592.

Detection probes are generally nucleic acid sequences or uncharged nucleic acid analogs such as, for example peptide nucleic acids which are disclosed in International Patent  
20 Application WO 92/20702, morpholino analogs which are described in U.S. Patents Numbered 5,185,444; 5,034,506 and 5,142,047. The probe may have to be rendered "non-extendable" in that additional dNTPs cannot be added to the probe. In and of themselves analogs usually are non-extendable and nucleic acid probes can be rendered non-extendable by modifying the 3' end of the probe such that the hydroxyl group is no longer capable of participating in elongation.  
25 For example, the 3' end of the probe can be functionalized with the capture or detection label to thereby consume or otherwise block the hydroxyl group. Alternatively, the 3' hydroxyl group simply can be cleaved, replaced or modified; U.S. Patent Application Serial No. 07/049,061 filed April 19, 1993, describes modifications which can be used to render a probe non-extendable.

30 Any of the polynucleotides of the present invention can be labeled, if desired, by incorporating a label detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive substances ( $^{32}\text{P}$ ,  $^{35}\text{S}$ ,  $^3\text{H}$ ,  $^{125}\text{I}$ ), fluorescent dyes (5-bromodesoxyuridin, fluorescein, acetylaminofluorene, digoxigenin) or biotin. Preferably, polynucleotides are labeled at their 3'  
35 and 5' ends. Examples of non-radioactive labeling of nucleic acid fragments are described in

the French patent No. FR-7810975 or by Urdea et al (1988) or Sanchez-Pescador et al (1988). In addition, the probes according to the present invention may have structural characteristics such that they allow the signal amplification, such structural characteristics being, for example, branched DNA probes as those described by Urdea et al. in 1991 or in the European patent No. EP 0 225 807 (Chiron).

A label can also be used to capture the primer, so as to facilitate the immobilization of either the primer or a primer extension product, such as amplified DNA, on a solid support. A capture label is attached to the primers or probes and can be a specific binding member which forms a binding pair with the solid's phase reagent's specific binding member (e.g. biotin and streptavidin). Therefore depending upon the type of label carried by a polynucleotide or a probe, it may be employed to capture or to detect the target DNA. Further, it will be understood that the polynucleotides, primers or probes provided herein, may, themselves, serve as the capture label. For example, in the case where a solid phase reagent's binding member is a nucleic acid sequence, it may be selected such that it binds a complementary portion of a primer or probe to thereby immobilize the primer or probe to the solid phase. In cases where a polynucleotide probe itself serves as the binding member, those skilled in the art will recognize that the probe will contain a sequence or "tail" that is not complementary to the target. In the case where a polynucleotide primer itself serves as the capture label, at least a portion of the primer will be free to hybridize with a nucleic acid on a solid phase. DNA Labeling techniques are well known to the skilled technician.

The probes of the present invention are useful for a number of purposes. They can be notably used in Southern hybridization to genomic DNA. The probes can also be used to detect PCR amplification products. They may also be used to detect mismatches in a sequence comprising a polynucleotide of SEQ ID Nos 1 to 17, 24, 25, 28, 74 and 79 to 132, or a g35030 polynucleotide or gene or mRNA using other techniques.

Any of the polynucleotides, primers and probes of the present invention can be conveniently immobilized on a solid support. Solid supports are known to those skilled in the art and include the walls of wells of a reaction tray, test tubes, polystyrene beads, magnetic beads, nitrocellulose strips, membranes, microparticles such as latex particles, sheep (or other animal) red blood cells, duracytes and others. The solid support is not critical and can be selected by one skilled in the art. Thus, latex particles, microparticles, magnetic or non-magnetic beads, membranes, plastic tubes, walls of microtiter wells, glass or silicon chips, sheep (or other suitable animal's) red blood cells and duracytes are all suitable examples. Suitable methods for immobilizing nucleic acids on solid phases include ionic, hydrophobic, covalent interactions and the like. A solid support, as used herein, refers to any material which



is insoluble, or can be made insoluble by a subsequent reaction. The solid support can be chosen for its intrinsic ability to attract and immobilize the capture reagent. Alternatively, the solid phase can retain an additional receptor which has the ability to attract and immobilize the capture reagent. The additional receptor can include a charged substance that is oppositely charged with respect to the capture reagent itself or to a charged substance conjugated to the capture reagent. As yet another alternative, the receptor molecule can be any specific binding member which is immobilized upon (attached to) the solid support and which has the ability to immobilize the capture reagent through a specific binding reaction. The receptor molecule enables the indirect binding of the capture reagent to a solid support material before the performance of the assay or during the performance of the assay. The solid phase thus can be a plastic, derivatized plastic, magnetic or non-magnetic metal, glass or silicon surface of a test tube, microtiter well, sheet, bead, microparticle, chip, sheep (or other suitable animal's) red blood cells, duracytes and other configurations known to those of ordinary skill in the art. The polynucleotides of the invention can be attached to or immobilized on a solid support individually or in groups of at least 2, 5, 8, 10, 12, 15, 20, or 25 distinct polynucleotides of the invention to a single solid support. In addition, polynucleotides other than those of the invention may be attached to the same solid support as one or more polynucleotides of the invention.

Consequently, the invention also comprises a method for detecting the presence of a nucleic acid comprising a nucleotide sequence selected from a group consisting of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132, a fragment or a variant thereof or a complementary sequence thereto in a sample, said method comprising the following steps of:

a) bringing into contact a nucleic acid probe or a plurality of nucleic acid probes which can hybridize with a nucleotide sequence included in a nucleic acid selected from the group consisting of the nucleotide sequences of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132, a fragment or a variant thereof or a complementary sequence thereto and the sample to be assayed; and

b) detecting the hybrid complex formed between the probe and a nucleic acid in the sample.

The invention further concerns a kit for detecting the presence of a nucleic acid comprising a nucleotide sequence selected from a group consisting of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132, a fragment or a variant thereof or a complementary sequence thereto in a sample, said kit comprising:

a) a nucleic acid probe or a plurality of nucleic acid probes which can hybridize with a nucleotide sequence included in a nucleic acid selected from the group consisting of the

nucleotide sequences of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132, a fragment or a variant thereof or a complementary sequence thereto; and

b) optionally, the reagents necessary for performing the hybridization reaction.

In a first preferred embodiment of this detection method and kit, said nucleic acid probe or the plurality of nucleic acid probes are labeled with a detectable molecule. In a second preferred embodiment of said method and kit, said nucleic acid probe or the plurality of nucleic acid probes has been immobilized on a substrate. In a third preferred embodiment, the nucleic acid probe or the plurality of nucleic acid probes comprise either a sequence which is selected from the group consisting of the nucleotide sequences of P1 to P127 and the complementary sequence thereto, B1 to B114, C1 to C114, D1 to D127, E1 to E127, or a nucleotide sequence comprising a biallelic marker selected from the group consisting of A1 to A127, or the complements thereto.

#### Oligonucleotide Arrays

A substrate comprising a plurality of oligonucleotide primers or probes of the invention may be used either for detecting or amplifying targeted sequences in a nucleotide sequence of SEQ ID No. 1, more particularly in a g35030 polynucleotide, or in genes comprising a g35030 polynucleotide and may also be used for detecting mutations in the coding or in the non-coding sequences of a g35030 nucleic acid sequence, or genes comprising a g35030 nucleic acid sequence.

Any polynucleotide provided herein may be attached in overlapping areas or at random locations on the solid support. Alternatively the polynucleotides of the invention may be attached in an ordered array wherein each polynucleotide is attached to a distinct region of the solid support which does not overlap with the attachment site of any other polynucleotide. Preferably, such an ordered array of polynucleotides is designed to be "addressable" where the distinct locations are recorded and can be accessed as part of an assay procedure. Addressable polynucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. The knowledge of the precise location of each polynucleotides location makes these "addressable" arrays particularly useful in hybridization assays. Any addressable array technology known in the art can be employed with the polynucleotides of the invention. One particular embodiment of these polynucleotide arrays is known as Genechips™, and has been generally described in US Patent 5,143,854; PCT publications WO 90/15070 and 92/10092. These arrays may generally be produced using mechanical synthesis methods or light directed synthesis methods which incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis (Fodor et al., 1991, incorporated herein by reference). The immobilization of arrays of oligonucleotides

on solid supports has been rendered possible by the development of a technology generally identified as "Very Large Scale Immobilized Polymer Synthesis" (VLSIPS™) in which, typically, probes are immobilized in a high density array on a solid surface of a chip. Examples of VLSIPS™ technologies are provided in US Patents 5,143,854; and 5,412,087 and in PCT Publications WO 90/15070, WO 92/10092 and WO 95/11995, which describe methods for forming oligonucleotide arrays through techniques such as light-directed synthesis techniques. In designing strategies aimed at providing arrays of nucleotides immobilized on solid supports, further presentation strategies were developed to order and display the oligonucleotide arrays on the chips in an attempt to maximize hybridization patterns and sequence information. Examples of such presentation strategies are disclosed in PCT Publications WO 94/12305, WO 94/11530, WO 97/29212 and WO 97/31256.

In another embodiment of the oligonucleotide arrays of the invention, an oligonucleotide probe matrix may advantageously be used to detect mutations occurring in a g35030 polynucleotide, including in genes comprising a g35030 polynucleotide and preferably in a g35030 polynucleotide regulatory region. For this particular purpose, probes are specifically designed to have a nucleotide sequence allowing their hybridization to the genes that carry known mutations (either by deletion, insertion or substitution of one or several nucleotides). By known mutations in a g35030 polynucleotide, it is meant, mutations in a g35030 polynucleotide that have been identified according; the technique used by Huang et al.(1996) or Samson et al.(1996), for example, may be used to identify such mutations.

Another technique that is used to detect mutations in a g35030 polynucleotide is the use of a high-density DNA array. Each oligonucleotide probe constituting a unit element of the high density DNA array is designed to match a specific subsequence of a g35030 polynucleotide. Thus, an array consisting of oligonucleotides complementary to subsequences of the target gene sequence is used to determine the identity of the target sequence with the wild-type gene sequence, measure its amount, and detect differences between the target sequence and the reference wild-type nucleic acid sequence of a g35030 polynucleotide. In one such design, termed 4L tiled array, is implemented a set of four probes (A, C, G, T), preferably 15-nucleotide oligomers. In each set of four probes, the perfect complement will hybridize more strongly than mismatched probes. Consequently, a nucleic acid target of length L is scanned for mutations with a tiled array containing 4L probes, the whole probe set containing all the possible mutations in the known wild reference sequence. The hybridization signals of the 15-mer probe set tiled array are perturbed by a single base change in the target sequence. As a consequence, there is a characteristic loss of signal or a "footprint" for the probes flanking a mutation position. This technique was described by Chee et al. in 1996.

Consequently, the invention concerns an array of nucleic acid molecules comprising at least one polynucleotide described above as probes and primers. Preferably, the invention concerns an array of nucleic acid comprising at least two polynucleotides described above as probes and primers.

5                   **G35030 Proteins and Polypeptide Fragments:**

The terms "g35030 polypeptides", are used herein to embrace all of the proteins and polypeptides encoded by the respective g35030 polypeptides of the present invention. Forming part of the invention are polypeptides encoded by the polynucleotides of the invention, as well as fusion polypeptides comprising such polypeptides. The invention embodies proteins from  
10 humans, mammals, primates, non-human primates, and includes isolated or purified g35030 proteins consisting, consisting essentially, or comprising the sequence of SEQ ID Nos 18 to 23.

It should be noted that the g35030 proteins of the invention also comprise naturally-occurring variants of the amino acid sequence of the respective human g35030 proteins.

The present invention embodies isolated, purified, and recombinant polypeptides  
15 comprising a contiguous span of at least 4 amino acids, preferably at least 6, more preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 95 or 100 amino acids, to the extent that said span is consistent with the length of a particular SEQ ID, of SEQ ID Nos 18 to 23. In other preferred embodiments the contiguous stretch of amino acids comprises the site of a mutation or functional mutation, including a deletion, addition,  
20 swap or truncation of the amino acids in a g35030 protein sequence. In one aspect, the contiguous stretch of amino acids comprises deletion, addition, swap or truncation at amino acid position 13 of SEQ ID Nos 18 to 23, corresponding to polymorphism described herein by biallelic markers 8-126-286 (also referred to as 8-130-143) and 8-155-258.

The invention further embodies g35030 polypeptides, including isolated and  
25 recombinant polypeptides, encoded respectively by g35030 polynucleotides consisting, consisting essentially, or comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or 500 nucleotides of SEQ ID Nos 2 to 17.

The invention further embodies g35030 polypeptides, including isolated and recombinant polypeptides, encoded respectively by g35030 polynucleotides consisting,  
30 consisting essentially, or comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or 500 nucleotides, to the extent that the length of said span is consistent with the nucleotide position range, of SEQ ID No 1, wherein said contiguous span comprises at least 1, 2, 3, 4, 5, 7 or 10 of the following nucleotide positions of SEQ ID No 1: 201123 to 201234, 201123 to 201560, 214676 to 214793, 215702 to 215746, 216836 to

216994, 216836 to 217077, 217671 to 217764, 227655 to 227736, 238715 to 238919, 240440 to 240673, 240440 to 241153, 241072 to 241291, 244353 to 244561 and 246273 to 247802, or the complements thereof.

g35030 proteins are preferably isolated from human or mammalian tissue samples or expressed from human or mammalian genes. The g35030 polypeptides of the invention can be made using routine expression methods known in the art. The polynucleotide encoding the desired polypeptide, is ligated into an expression vector suitable for any convenient host. Both eukaryotic and prokaryotic host systems is used in forming recombinant polypeptides, and a summary of some of the more common systems. The polypeptide is then isolated from lysed cells or from the culture medium and purified to the extent needed for its intended use. Purification is by any technique known in the art, for example, differential extraction, salt fractionation, chromatography, centrifugation, and the like. See, for example, Methods in Enzymology for a variety of methods for purifying proteins.

In addition, shorter protein fragments can be produced by chemical synthesis. Alternatively the proteins of the invention is extracted from cells or tissues of humans or non-human animals. Methods for purifying proteins are known in the art, and include the use of detergents or chaotropic agents to disrupt particles followed by differential extraction and separation of the polypeptides by ion exchange chromatography, affinity chromatography, sedimentation according to density, and gel electrophoresis.

Any g35030 cDNA or fragment thereof, including the respective cDNA sequences of SEQ ID Nos 2 to 17, is used to express g35030 proteins and polypeptides. The nucleic acid encoding the g35030 protein or polypeptide to be expressed is operably linked to a promoter in an expression vector using conventional cloning technology. The g35030 insert in the expression vector may comprise the full coding sequence for the respective g35030 protein or a portion thereof. For example, the g35030 derived insert may encode a polypeptide comprising at least 10 consecutive amino acids of the respective g35030 protein of SEQ ID Nos 18 to 23.

The expression vector is any of the mammalian, yeast, insect or bacterial expression systems known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Genetics Institute (Cambridge, MA), Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence is optimized for the particular expression organism in which the expression vector is introduced, as explained by Hatfield, et al., U.S. Patent No. 5,082,767.

In one embodiment, the entire coding sequence of the g35030 cDNA through the poly A signal of the cDNA are operably linked to a promoter in the expression vector. Alternatively, if the

nucleic acid encoding a portion of the g35030 protein lacks a methionine to serve as the initiation site, an initiating methionine can be introduced next to the first codon of the nucleic acid using conventional techniques. Similarly, if the insert from the g35030 cDNA lacks a poly A signal, this sequence can be added to the construct by, for example, splicing out the Poly A signal from pSG5 (Stratagene) using BglI and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene). pXT1 contains the LTRs and a portion of the gag gene from Moloney Murine Leukemia Virus. The position of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex Thymidine Kinase promoter and the selectable neomycin gene. The nucleic acid encoding the g35030 protein or a portion thereof is obtained by PCR from a bacterial vector containing the a nucleotide sequence of an exon of a g35030 gene as described herein and in SEQ ID No 1, or from a g35030 cDNA comprising a nucleic acid of SEQ ID No 2 to 17 using oligonucleotide primers complementary to the g35030 nucleic acid or portion thereof and containing restriction endonuclease sequences for Pst I incorporated into the 5' primer and BglII at the 5' end of the corresponding cDNA 3' primer, taking care to ensure that the sequence encoding the g35030 protein or a portion thereof is positioned properly with respect to the poly A signal. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with Bgl II, purified and ligated to pXT1, now containing a poly A signal and digested with BglII.

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600ug/ml G418 (Sigma, St. Louis, Missouri).

Alternatively, the nucleic acids encoding the g35030 protein or a portion thereof is cloned into pED6dpc2 (Genetics Institute, Cambridge, MA). The resulting pED6dpc2 constructs is transfected into a suitable host cell, such as COS 1 cells. Methotrexate resistant cells are selected and expanded.

The above procedures may also be used to express a mutant g35030 protein responsible for a detectable phenotype or a portion thereof.

The expressed proteins are purified using conventional purification techniques such as ammonium sulfate precipitation or chromatographic separation based on size or charge. The protein encoded by the nucleic acid insert may also be purified using standard immunochromatography techniques. In such procedures, a solution containing the expressed g35030 protein or portion thereof, such as a cell extract, is applied to a column having antibodies against the g35030 protein or portion thereof is attached to the chromatography matrix. The expressed protein is allowed to bind the immunochromatography column. Thereafter, the column

is washed to remove non-specifically bound proteins. The specifically bound expressed protein is then released from the column and recovered using standard techniques.

To confirm expression of the g35030 protein or a portion thereof, the proteins expressed from host cells containing an expression vector containing an insert encoding the g35030 protein or a portion thereof can be compared to the proteins expressed in host cells containing the expression vector without an insert. The presence of a band in samples from cells containing the expression vector with an insert which is absent in samples from cells containing the expression vector without an insert indicates that the g35030 protein or a portion thereof is being expressed. Generally, the band will have the mobility expected for the g35030 protein or portion thereof. However, the band may have a mobility different than that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage.

Antibodies capable of specifically recognizing the expressed g35030 protein or a portion thereof are described below.

If antibody production is not possible, the nucleic acids encoding the g35030 protein or a portion thereof is incorporated into expression vectors designed for use in purification schemes employing chimeric polypeptides. In such strategies the nucleic acid encoding the g35030 protein or a portion thereof is inserted in frame with the gene encoding the other half of the chimera. The other half of the chimera is  $\beta$ -globin or a nickel binding polypeptide encoding sequence. A chromatography matrix having antibody to  $\beta$ -globin or nickel attached thereto is then used to purify the chimeric protein. Protease cleavage sites is engineered between the  $\beta$ -globin gene or the nickel binding polypeptide and the g35030 protein or portion thereof. Thus, the two polypeptides of the chimera is separated from one another by protease digestion.

One useful expression vector for generating  $\beta$ -globin chimeric proteins is pSG5 (Stratagene), which encodes rabbit  $\beta$ -globin. Intron II of the rabbit  $\beta$ -globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis et al., (1986) and many of the methods are available from Stratagene, Life Technologies, Inc., or Promega. Polypeptide may additionally be produced from the construct using in vitro translation systems such as the In vitro Express<sup>TM</sup> Translation Kit (Stratagene).

#### **Antibodies That Bind g35030 Polypeptides of the Invention**

Any g35030 polypeptide or whole protein may be used to generate antibodies capable of specifically binding to an expressed g35030 protein or fragments thereof.

For an antibody composition to specifically bind to a g35030 protein, it must

demonstrate at least a 5%, 10%, 15%, 20%, 25%, 50%, or 100% greater binding affinity for full length g35030 protein than for any full length protein in an ELISA, RIA, or other antibody-based binding assay. For an antibody composition to specifically bind to a variant g35030 protein, it must demonstrate at least a 5%, 10%, 15%, 20%, 25%, 50%, or 100% greater binding  
5 affinity for the respective full length variant g35030 protein than for the respective reference g35030 full length protein in an ELISA, RIA, or other antibody-based binding assay.

One antibody composition of the invention is capable of specifically binding or specifically binds to the respective g35030 proteins of SEQ ID Nos 18 to 23. Other antibody compositions of the invention are capable of specifically binding or specifically bind to a  
10 g35030 protein variant.

In one embodiment, the invention concerns antibody compositions, either polyclonal or monoclonal, capable of selectively binding, or selectively bind to an epitope-containing a polypeptide comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a g35030  
15 polypeptide.

The invention also concerns a purified or isolated antibody capable of specifically binding to a mutated g35030 protein or to a fragment or variant thereof comprising an epitope of the mutated g35030 protein. In another preferred embodiment, the present invention concerns an antibody capable of binding to a polypeptide comprising at least 10 consecutive  
20 amino acids of a g35030 protein and including at least one of the amino acids which can be encoded by the trait causing mutations.

An epitope can comprise as few as 3 amino acids in a spatial conformation, which is unique to the epitope. Generally an epitope consists of at least 6 such amino acids, and more often at least 8-10 such amino acids. In preferred embodiment, antigenic epitopes comprise a  
25 number of amino acids that is any integer between 3 and 50. Fragments which function as epitopes may be produced by any conventional means. Epitopes can be determined by a Jameson-Wolf antigenic analysis, for example, performed using the computer program PROTEAN, using default parameters (Version 4.0 Windows, DNASTAR, Inc., 1228 South Park Street Madison, WI).

Predicted antigenic epitopes are shown below. It is pointed out that the immunogenic epitope list describe only amino acid residues comprising epitopes predicted to have the highest degree of immunogenicity by a particular algorithm. Polypeptides of the present invention that are not specifically described as immunogenic are not considered non-antigenic. This is  
30 because they may still be antigenic *in vivo* but merely not recognized as such by the particular algorithm used. Alternatively, the polypeptides are probably antigenic *in vitro* using methods  
35



such a phage display. Thus, listed below are the amino acid residues comprising only preferred epitopes, not a complete list. In fact, all fragments of the polypeptides of the present invention, at least 6 amino acids residues in length, are included in the present invention as being useful as antigenic epitope. Moreover, listed below are only the critical residues of the epitopes  
5 determined by the Jameson-Wolf analysis. Thus, additional flanking residues on either the N-terminal, C-terminal, or both N- and C-terminal ends may be added to the sequences listed to generate an epitope-bearing portion at least 6 residues in length. Amino acid residues comprising other immunogenic epitopes may be determined by algorithms similar to the Jameson-Wolf analysis or by *in vivo* testing for an antigenic response using the methods  
10 described herein or those known in the art.

The epitope-bearing fragments of the present invention preferably comprises 6 to 50 amino acids (i.e. any integer between 6 and 50, inclusive) of a polypeptide of the present invention. Also, included in the present invention are antigenic fragments between the integers of 6 and the full length g35030 sequence.

15 Preferred g35030 immunogenic epitopes:

SEQ ID No. 18: Val-9 to Arg-11; Gly-21 to Lys-22

SEQ ID No. 19: Val-9 to Arg-11; Pro-45 to Pro-52

SEQ ID No. 20: Val-9 to Arg-11

SEQ ID No. 21: Val-9 to Arg-11; Ser-40 to Ser-42

20 SEQ ID No. 22: Val-9 to Arg-11

SEQ ID No. 23: Val-9 to Arg-11; Pro-45 to Glu-49; Ser-57 to Val-60

Non-human animals or mammals, whether wild-type or transgenic, which express a different species of g35030 than the one to which antibody binding is desired, and animals  
25 which do not express g35030 (i.e. a g35030 knock out animal as described herein) are particularly useful for preparing antibodies. g35030 knock out animals will recognize all or most of the exposed regions of a g35030 protein as foreign antigens, and therefore produce antibodies with a wider array of g35030 epitopes. Moreover, smaller polypeptides with only 10 to 30 amino acids may be useful in obtaining specific binding to any one of the g35030 proteins.  
30 In addition, the humoral immune system of animals which produce a species of g35030 that resembles the antigenic sequence will preferentially recognize the differences between the animal's native g35030 species and the antigen sequence, and produce antibodies to these unique sites in the antigen sequence. Such a technique will be particularly useful in obtaining antibodies that specifically bind to any one of the g35030 proteins.

35 Antibody preparations prepared according to either protocol are useful in quantitative

immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

5           The antibodies of the invention may be labeled by any one of the radioactive, fluorescent or enzymatic labels known in the art.

Consequently, the invention is also directed to a method for detecting specifically the presence of a g35030 polypeptide according to the invention in a biological sample, said method comprising the following steps:

- 10           a) bringing into contact the biological sample with a polyclonal or monoclonal antibody that specifically binds a g35030 polypeptide comprising an amino acid sequence of SEQ ID Nos. 18 to 23, or to a peptide fragment or variant thereof; and
- b) detecting the antigen-antibody complex formed.

15           The invention also concerns a diagnostic kit for detecting *in vitro* the presence of a g35030 polypeptide according to the present invention in a biological sample, wherein said kit comprises:

- a) a polyclonal or monoclonal antibody that specifically binds a g35030 polypeptide comprising an amino acid sequence of SEQ ID Nos 18 to 23, or to a peptide fragment or variant thereof, optionally labeled;
- 20           b) a reagent allowing the detection of the antigen-antibody complexes formed, said reagent carrying optionally a label, or being able to be recognized itself by a labeled reagent, more particularly in the case when the above-mentioned monoclonal or polyclonal antibody is not labeled by itself.

25           The present invention thus relates to antibodies and T-cell antigen receptors (TCR), which specifically bind the polypeptides, and more specifically, the epitopes of the polypeptides of the present invention, including but not limited to IgG (including IgG1, IgG2, IgG3, and IgG4), IgA (including IgA1 and IgA2), IgD, IgE, or IgM, and IgY. In a preferred embodiment the antibodies are human antigen binding antibody fragments of the present invention include, but are not limited to, Fab, Fab' F(ab)2 and F(ab')2, Fd, single-chain Fvs

30           (scFv), single-chain antibodies, disulfide-linked Fvs (sdFv) and fragments comprising either a V<sub>L</sub> or V<sub>H</sub> domain. The antibodies may be from any animal origin including birds and mammals. Preferably, the antibodies are human, murine, rabbit, goat, guinea pig, camel, horse, or chicken.

35           Antigen-binding antibody fragments, including single-chain antibodies, may comprise the variable region(s) alone or in combination with the entire or partial of the following: hinge region, CH1, CH2, and CH3 domains. Also included in the invention are any combinations of

variable region(s) and hinge region, CH1, CH2, and CH3 domains. The present invention further includes chimeric, humanized, and human monoclonal and polyclonal antibodies, which specifically bind the polypeptides of the present invention. The present invention further includes antibodies that are anti-idiotypic to the antibodies of the present invention.

5           The antibodies of the present invention may be monospecific, bispecific, trispecific or have greater multispecificity. Multispecific antibodies may be specific for different epitopes of a polypeptide of the present invention or may be specific for both a polypeptide of the present invention as well as for heterologous compositions, such as a heterologous polypeptide or solid support material. See, e.g., WO 93/17715; WO 92/08802; WO 91/00360; WO 92/05793; Tutt, 10   A. et al. (1991) *J. Immunol.* 147:60-69; US Patents 5,573,920, 4,474,893, 5,601,819, 4,714,681, 4,925,648; Kostelny, S.A. et al. (1992) *J. Immunol.* 148:1547-1553.

          Antibodies of the present invention may be described or specified in terms of the epitope(s) or epitope-bearing portion(s) of a polypeptide of the present invention, which are recognized or specifically bound by the antibody. In the case of proteins of the present 15   invention secreted proteins, the antibodies may specifically bind a full-length protein encoded by a nucleic acid of the present invention, a mature protein (i.e., the protein generated by cleavage of the signal peptide) encoded by a nucleic acid of the present invention, a signal peptide encoded by a nucleic acid of the present invention, or any other polypeptide of the present invention. Therefore, the epitope(s) or epitope bearing polypeptide portion(s) may be 20   specified as described herein, e.g., by N-terminal and C-terminal positions, by size in contiguous amino acid residues, or otherwise described herein (including the sequence listing). Antibodies which specifically bind any epitope or polypeptide of the present invention may also be excluded as individual species. Therefore, the present invention includes antibodies that specifically bind specified polypeptides of the present invention, and allows for the exclusion of 25   the same.

          Antibodies of the present invention may also be described or specified in terms of their cross-reactivity. Antibodies that do not specifically bind any other analog, ortholog, or homolog of the polypeptides of the present invention are included. Antibodies that do not bind polypeptides with less than 95%, less than 90%, less than 85%, less than 80%, less than 75%, 30   less than 70%, less than 65%, less than 60%, less than 55%, and less than 50% identity (as calculated using methods known in the art and described herein) to a polypeptide of the present invention are also included in the present invention. Further included in the present invention are antibodies, which only bind polypeptides encoded by polynucleotides, which hybridize to a polynucleotide of the present invention under stringent hybridization conditions (as described 35   herein). Antibodies of the present invention may also be described or specified in terms of their

binding affinity. Preferred binding affinities include those with a dissociation constant or  $K_d$  less than  $5 \times 10^{-6}M$ ,  $10^{-6}M$ ,  $5 \times 10^{-7}M$ ,  $10^{-7}M$ ,  $5 \times 10^{-8}M$ ,  $10^{-8}M$ ,  $5 \times 10^{-9}M$ ,  $10^{-9}M$ ,  $5 \times 10^{-10}M$ ,  $10^{-10}M$ ,  $5 \times 10^{-11}M$ ,  $10^{-11}M$ ,  $5 \times 10^{-12}M$ ,  $10^{-12}M$ ,  $5 \times 10^{-13}M$ ,  $10^{-13}M$ ,  $5 \times 10^{-14}M$ ,  $10^{-14}M$ ,  $5 \times 10^{-15}M$ , and  $10^{-15}M$ .

5           Antibodies of the present invention have uses that include, but are not limited to, methods known in the art to purify, detect, and target the polypeptides of the present invention including both in vitro and in vivo diagnostic and therapeutic methods. For example, the antibodies have use in immunoassays for qualitatively and quantitatively measuring levels of the polypeptides of the present invention in biological samples. See, e.g., Harlow et al.,  
10   ANTIBODIES: A LABORATORY MANUAL, (Cold Spring Harbor Laboratory Press, 2nd ed. 1988) (incorporated by reference in the entirety).

          The antibodies of the present invention may be used either alone or in combination with other compositions. The antibodies may further be recombinantly fused to a heterologous polypeptide at the N- or C-terminus or chemically conjugated (including covalent and non-  
15   covalent conjugations) to polypeptides or other compositions. For example, antibodies of the present invention may be recombinantly fused or conjugated to molecules useful as labels in detection assays and effector molecules such as heterologous polypeptides, drugs, or toxins. See, e.g., WO 92/08495; WO 91/14438; WO 89/12624; US Patent 5,314,995; and EP 0 396 387.

          The antibodies of the present invention may be prepared by any suitable method known  
20   in the art. For example, a polypeptide of the present invention or an antigenic fragment thereof can be administered to an animal in order to induce the production of sera containing polyclonal antibodies. The term "monoclonal antibody" is not limited to antibodies produced through hybridoma technology. The term "antibody" refers to a polypeptide or group of polypeptides which are comprised of at least one binding domain, where a binding domain is formed from  
25   the folding of variable domains of an antibody molecule to form three-dimensional binding spaces with an internal surface shape and charge distribution complementary to the features of an antigenic determinant of an antigen., which allows an immunological reaction with the antigen. The term "monoclonal antibody" refers to an antibody that is derived from a single clone, including eukaryotic, prokaryotic, or phage clone, and not the method by which it is  
30   produced. Monoclonal antibodies can be prepared using a wide variety of techniques known in the art including the use of hybridoma, recombinant, and phage display technology.

          Hybridoma techniques include those known in the art (See, e.g., Harlow et al. (1998); Hammerling, et al. (1981) (said references incorporated by reference in their entireties). Fab and  $F(ab')_2$  fragments may be produced, for example, from hybridoma-produced antibodies by  
35   proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to

produce F(ab')<sub>2</sub> fragments).

Alternatively, antibodies of the present invention can be produced through the application of recombinant DNA technology or through synthetic chemistry using methods known in the art. For example, the antibodies of the present invention can be prepared using various phage display methods known in the art. In phage display methods, functional antibody domains are displayed on the surface of a phage particle, which carries polynucleotide sequences encoding them. Phage with a desired binding property are selected from a repertoire or combinatorial antibody library (e.g. human or murine) by selecting directly with antigen, typically antigen bound or captured to a solid surface or bead. Phage used in these methods are typically filamentous phage including fd and M13 with Fab, Fv or disulfide stabilized Fv antibody domains recombinantly fused to either the phage gene III or gene VIII protein. Examples of phage display methods that can be used to make the antibodies of the present invention include those disclosed in Brinkman U. et al. (1995); Ames, R.S. et al. (1995); Kettleborough, C.A. et al. (1994); Persic, L. et al. (1997); Burton, D.R. et al. (1994); PCT/GB91/01134; WO 90/02809; WO 91/10737; WO 92/01047; WO 92/18619; WO 93/11236; WO 95/15982; WO 95/20401; and US Patents 5,698,426, 5,223,409, 5,403,484, 5,580,717, 5,427,908, 5,750,753, 5,821,047, 5,571,698, 5,427,908, 5,516,637, 5,780,225, 5,658,727 and 5,733,743 (said references incorporated by reference in their entireties).

As described in the above references, after phage selection, the antibody coding regions from the phage can be isolated and used to generate whole antibodies, including human antibodies, or any other desired antigen binding fragment, and expressed in any desired host including mammalian cells, insect cells, plant cells, yeast, and bacteria. For example, techniques to recombinantly produce Fab, Fab' F(ab')<sub>2</sub> and F(ab')<sub>2</sub> fragments can also be employed using methods known in the art such as those disclosed in WO 92/22324; Mullinax, R.L. et al. (1992); and Sawai, H. et al. (1995); and Better, M. et al. (1988) (said references incorporated by reference in their entireties).

Examples of techniques which can be used to produce single-chain Fvs and antibodies include those described in U.S. Patents 4,946,778 and 5,258,498; Huston et al. (1991); Shu, L. et al. (1993); and Skerra, A. et al. (1988). For some uses, including in vivo use of antibodies in humans and in vitro detection assays, it may be preferable to use chimeric, humanized, or human antibodies. Methods for producing chimeric antibodies are known in the art. See e.g., Morrison (1985); Oi et al., (1986); Gillies, S.D. et al. (1989); and US Patent 5,807,715. Antibodies can be humanized using a variety of techniques including CDR-grafting (EP 0 239 400; WO 91/09967; US Patent 5,530,101; and 5,585,089), veneering or resurfacing (EP 0 592 106; EP 0 519 596; Padlan E.A., (1991); Studnicka G.M. et al. (1994); Roguska M.A. et al.

(1994), and chain shuffling (US Patent 5,565,332). Human antibodies can be made by a variety of methods known in the art including phage display methods described above. See also, US Patents 4,444,887, 4,716,111, 5,545,806, and 5,814,318; WO 98/46645; WO 98/50433; WO 98/24893; WO 96/34096; WO 96/33735; and WO 91/10741 (said references incorporated by reference in their entirety).

Further included in the present invention are antibodies recombinantly fused or chemically conjugated (including both covalently and non-covalently conjugations) to a polypeptide of the present invention. The antibodies may be specific for antigens other than polypeptides of the present invention. For example, antibodies may be used to target the polypeptides of the present invention to particular cell types, either in vitro or in vivo, by fusing or conjugating the polypeptides of the present invention to antibodies specific for particular cell surface receptors. Antibodies fused or conjugated to the polypeptides of the present invention may also be used in in vitro immunoassays and purification methods using methods known in the art. See e.g., Harbor et al. supra and WO 93/21232; EP 0 439 095; Naramura, M. et al. (1994); US Patent 5,474,981; Gillies, S.O. et al. (1992); Fell, H.P. et al. (1991) (said references incorporated by reference in their entirety).

The present invention further includes compositions comprising the polypeptides of the present invention fused or conjugated to antibody domains other than the variable regions. For example, the polypeptides of the present invention may be fused or conjugated to an antibody Fc region, or portion thereof. The antibody portion fused to a polypeptide of the present invention may comprise the hinge region, CH1 domain, CH2 domain, and CH3 domain or any combination of whole domains or portions thereof. The polypeptides of the present invention may be fused or conjugated to the above antibody portions to increase the in vivo half-life of the polypeptides or for use in immunoassays using methods known in the art. The polypeptides may also be fused or conjugated to the above antibody portions to form multimers. For example, Fc portions fused to the polypeptides of the present invention can form dimers through disulfide bonding between the Fc portions. Higher multimeric forms can be made by fusing the polypeptides to portions of IgA and IgM. Methods for fusing or conjugating the polypeptides of the present invention to antibody portions are known in the art. See e.g., US Patents 5,336,603, 5,622,929, 5,359,046, 5,349,053, 5,447,851, 5,112,946; EP 0 307 434, EP 0 367 166; WO 96/04388, WO 91/06570; Ashkenazi, A. et al. (1991); Zheng, X.X. et al. (1995); and Vil, H. et al. (1992) (said references incorporated by reference in their entirety).

The invention further relates to antibodies that act as agonists or antagonists of the polypeptides of the present invention. For example, the present invention includes antibodies that disrupt the receptor/ligand interactions with the polypeptides of the invention either

partially or fully. Included are both receptor-specific antibodies and ligand-specific antibodies. Included are receptor-specific antibodies, which do not prevent ligand binding but prevent receptor activation. Receptor activation (i.e., signaling) may be determined by techniques described herein or otherwise known in the art. Also include are receptor-specific antibodies which both prevent ligand binding and receptor activation. Likewise, included are neutralizing antibodies that bind the ligand and prevent binding of the ligand to the receptor, as well as antibodies that bind the ligand, thereby preventing receptor activation, but do not prevent the ligand from binding the receptor. Further included are antibodies that activate the receptor. These antibodies may act as agonists for either all or less than all of the biological activities affected by ligand-mediated receptor activation. The antibodies may be specified as agonists or antagonists for biological activities comprising specific activities disclosed herein. The above antibody agonists can be made using methods known in the art. See e.g., WO 96/40281; US Patent 5,811,097; Deng, B. et al. (1998); Chen, Z. et al. (1998); Harrop, J.A. et al. (1998); Zhu, Z. et al. (1998); Yoon, D.Y. et al. (1998); Prat, M. et al. (1998); Pitard, V. et al. (1997); Liautard, J. et al. (1997); Carlson, N.G. et al. (1997); Taryman, R.E. et al. (1995); Muller, Y.A. et al. (1998); Bartunek, P. et al. (1996) (said references incorporated by reference in their entireties).

As discussed above, antibodies of the polypeptides of the invention can, in turn, be utilized to generate anti-idiotypic antibodies that "mimic" polypeptides of the invention using techniques well known to those skilled in the art. See, e.g. Greenspan and Bona, (1989); Nissinoff, (1991). For example, antibodies which bind to and competitively inhibit polypeptide multimerization or binding of a polypeptide of the invention to ligand can be used to generate anti-idiotypes that "mimic" the polypeptide multimerization or binding domain and, as a consequence, bind to and neutralize polypeptide or its ligand. Such neutralization anti-idiotypic antibodies can be used to bind a polypeptide of the invention or to bind its ligands/receptors, and thereby block its biological activity.

#### **Biallelic markers of the inventions**

##### **Advantages of the biallelic markers of the present invention**

The biallelic marker of the inventions of the present invention offer a number of important advantages over other genetic markers such as RFLP (Restriction fragment length polymorphism) and VNTR (Variable Number of Tandem Repeats) markers.

The first generation of markers, were RFLPs, which are variations that modify the length of a restriction fragment. But methods used to identify and to type RFLPs are relatively wasteful of materials, effort, and time. The second generation of genetic markers were VNTRs,

which can be categorized as either minisatellites or microsatellites. Minisatellites are tandemly repeated DNA sequences present in units of 5-50 repeats which are distributed along regions of the human chromosomes ranging from 0.1 to 20 kilobases in length. Since they present many possible alleles, their informative content is very high. Minisatellites are scored by performing Southern blots to identify the number of tandem repeats present in a nucleic acid sample from the individual being tested. However, there are only  $10^4$  potential VNTRs that can be typed by Southern blotting. Moreover, both RFLP and VNTR markers are costly and time-consuming to develop and assay in large numbers.

Single nucleotide polymorphism or biallelic markers can be used in the same manner as RFLPs and VNTRs but offer several advantages. Single nucleotide polymorphisms are densely spaced in the human genome and represent the most frequent type of variation. An estimated number of more than  $10^7$  sites are scattered along the  $3 \times 10^9$  base pairs of the human genome. Therefore, single nucleotide polymorphism occur at a greater frequency and with greater uniformity than RFLP or VNTR markers which means that there is a greater probability that such a marker will be found in close proximity to a genetic locus of interest. Single nucleotide polymorphisms are less variable than VNTR markers but are mutationally more stable.

Also, the different forms of a characterized single nucleotide polymorphism, such as the biallelic markers of the present invention, are often easier to distinguish and can therefore be typed easily on a routine basis. Biallelic markers have single nucleotide based alleles and they have only two common alleles, which allows highly parallel detection and automated scoring. The biallelic markers of the present invention offer the possibility of rapid, high-throughput genotyping of a large number of individuals.

Biallelic markers are densely spaced in the genome, sufficiently informative and can be assayed in large numbers. The combined effects of these advantages make biallelic markers extremely valuable in genetic studies. Biallelic markers can be used in linkage studies in families, in allele sharing methods, in linkage disequilibrium studies in populations, in association studies of case-control populations. An important aspect of the present invention is that biallelic markers allow association studies to be performed to identify genes involved in complex traits. Association studies examine the frequency of marker alleles in unrelated case- and control-populations and are generally employed in the detection of polygenic or sporadic traits. Association studies may be conducted within the general population and are not limited to studies performed on related individuals in affected families (linkage studies). Biallelic markers in different genes can be screened in parallel for direct association with disease or response to a treatment. This multiple gene approach is a powerful tool for a variety of human genetic studies as it provides the necessary statistical power to examine the synergistic effect of



multiple genetic factors on a particular phenotype, drug response, sporadic trait, or disease state with a complex genetic etiology.

### **Biallelic Markers And Polynucleotides Comprising Them**

#### **5 Polynucleotides of the present invention**

In one aspect, the invention concerns biallelic markers associated with schizophrenia. The invention comprises chromosome 13q31-q33-related biallelic markers and g35030-related biallelic markers. The markers are generally referred to herein as A1, A2, A3 and so on. The biallelic markers of the invention comprise the biallelic markers designated A1 to A127 in  
10 Table 5. Also included are biallelic markers in linkage disequilibrium with the biallelic markers of the invention.

The polynucleotide of the invention may consist of, consist essentially of, or comprise a contiguous span of nucleotides of a sequence from any of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 as well as sequences which are complementary thereto ("complements thereof").  
15 The "contiguous span" may be at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500, 1000 or 2000 nucleotides in length, to the extent that a contiguous span of these lengths is consistent with the lengths of the particular Sequence ID.

The present invention encompasses polynucleotides for use as primers and probes in the methods of the invention. These polynucleotides may consist of, consist essentially of, or  
20 comprise a contiguous span of nucleotides of a sequence from any of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 as well as sequences which are complementary thereto ("complements thereof"). The "contiguous span" may be at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500, 1000 or 2000 nucleotides in length, to the extent that a contiguous span of these lengths is consistent with the lengths of the particular Sequence ID. It should be noted that the  
25 polynucleotides of the present invention are not limited to having the exact flanking sequences surrounding the polymorphic bases which, are enumerated in the Sequence Listing. Rather, it will be appreciated that the flanking sequences surrounding the biallelic markers and other polymorphisms of the invention, or any of the primers or probes of the invention which, are more distant from the markers, may be lengthened or shortened to any extent compatible with  
30 their intended use and the present invention specifically contemplates such sequences. It will be appreciated that the polynucleotides of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 may be of any length compatible with their intended use. Also the flanking regions outside of the contiguous span need not be homologous to native flanking sequences which actually occur in human subjects. The addition of any nucleotide sequence, which is compatible with the  
35 nucleotides intended use is specifically contemplated. The contiguous span may optionally

include the biallelic markers of the invention in said sequence. Biallelic markers generally comprise a polymorphism at one single base position. Each biallelic marker therefore corresponds to two forms of a polynucleotide sequence which, when compared with one another, present a nucleotide modification at one position. Usually, the nucleotide modification involves the substitution of one nucleotide for another. Optionally allele 1 or allele 2 of the biallelic markers disclosed in Table 5 may be specified as being present at the biallelic marker of the invention. Preferred polynucleotides may consist of, consist essentially of, or comprise a contiguous span of nucleotides of a sequence from SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 as well as sequences which are complementary thereto. The "contiguous span" may be at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500, 1000 or 2000 nucleotides in length, to the extent that a contiguous span of these lengths is consistent with the lengths of the particular Sequence ID.

A preferred probe or primer comprises a nucleic acid comprising a polynucleotide selected from the group of the nucleotide sequences of P1 to P127 and the complementary sequence thereto, B1 to B114, C1 to C114, D1 to D127, E1 to E127.

The invention also relates to polynucleotides that hybridize, under conditions of high or intermediate stringency, to a polynucleotide of any of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 as well as sequences, which are complementary thereto. Preferably such polynucleotides are at least 20, 25, 35, 40, 50, 70, 80, 100, 250, 500, 1000 or 2000 nucleotides in length, to the extent that a polynucleotide of these lengths is consistent with the lengths of the particular Sequence ID. Preferred polynucleotides comprise a polymorphism of the invention. Optionally either allele 1 or allele 2 of the polymorphism disclosed in Table 5 may be specified as being present at the polymorphism of the invention. Particularly preferred polynucleotides comprise a biallelic marker of the invention. Conditions of high stringency are further described herein.

The primers of the present invention may be designed from the disclosed sequences for any method known in the art. A preferred set of primers is fashioned such that the 3' end of the contiguous span of identity with the sequences of any of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 is present at the 3' end of the primer. Such a configuration allows the 3' end of the primer to hybridize to a selected nucleic acid sequence and dramatically increases the efficiency of the primer for amplification or sequencing reactions. In a preferred set of primers the contiguous span is found in one of the sequences described in Table 4. Allele specific primers may be designed such that a biallelic marker or other polymorphism of the invention is at the 3' end of the contiguous span and the contiguous span is present at the 3' end of the primer. Such allele specific primers tend to selectively prime an amplification or sequencing

reaction so long as they are used with a nucleic acid sample that contains one of the two alleles present at said marker. The 3' end of primer of the invention may be located within or at least 2, 4, 6, 8, 10, 12, 15, 18, 20, 25, 50, 100, 250, 500, or 1000 nucleotides upstream of a biallelic marker of the invention in said sequence or at any other location which is appropriate for their intended use in sequencing, amplification or the location of novel sequences or markers.

5 Primers with their 3' ends located 1 nucleotide upstream of an biallelic marker of the invention have a special utility as microsequencing assays. Preferred microsequencing primers are described in Table 6.

The probes of the present invention may be designed from the disclosed sequences for any method known in the art, particularly methods which allow for testing if a particular sequence or marker disclosed herein is present. A preferred set of probes may be designed for use in the hybridization assays of the invention in any manner known in the art such that they selectively bind to one allele of a biallelic marker or other polymorphism, but not the other under any particular set of assay conditions. Preferred hybridization probes may consists of,

10 consist essentially of, or comprise a contiguous span which ranges in length from 8, 10, 12, 15, 18 or 20 to 25, 35, 40, 50, 60, 70, or 80 nucleotides, or be specified as being 12, 15, 18, 20, 25, 35, 40, or 50 nucleotides in length and including an biallelic marker or other polymorphism of the invention in said sequence. In a preferred embodiment, either of allele 1 or 2 disclosed in Table 5 may be specified as being present at the biallelic marker site. In another preferred

20 embodiment, said biallelic marker may be within 6, 5, 4, 3, 2, or 1 nucleotides of the center of the hybridization probe or at the center of said probe.

In one embodiment the invention encompasses isolated, purified, and recombinant polynucleotides comprising, consisting of, or consisting essentially of a contiguous span of 8 to 50 nucleotides of any one of SEQ ID Nos 1 to 17, 24, 25, 28, 74 and 79 to 132 and the

25 complement thereof, wherein said span includes a biallelic marker of the invention, a chromosome 13q31-q33-related biallelic marker or g35030-related biallelic marker in said sequence; optionally, wherein said polymorphism or biallelic marker selected from the group consisting of A1 to A127, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said biallelic marker is selected from the

30 group consisting of A13 to A18, A20 to A46, A49 to A52, A55, A57, A59 to A63, A72 to A73, A76 and A123 ; optionally, wherein said biallelic marker is selected from the group consisting of A13 to A65, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said contiguous span is 18 to 35 nucleotides in length and said biallelic marker is within 4 nucleotides of the center of said polynucleotide;

35 optionally, wherein said polynucleotide consists of said contiguous span and said contiguous

span is 25 nucleotides in length and said biallelic marker is at the center of said polynucleotide; optionally, wherein the 3' end of said contiguous span is present at the 3' end of said polynucleotide; and optionally, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide and said biallelic marker is present at the 3' end of said polynucleotide. In a preferred embodiment, said probes comprise, consists of, or consists essentially of a sequence selected from the following sequences: P1 to P127 and the complementary sequences thereto.

In another embodiment the invention encompasses isolated, purified and recombinant polynucleotides comprising, consisting of, or consisting essentially of a contiguous span of 8 to 50 nucleotides of any one of SEQ ID Nos 1 to 17, 24, 25, 28, 74 and 79 to 132, or the complement thereof, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide, and wherein the 3' end of said polynucleotide is located within 20 nucleotides upstream of a polymorphism of the invention, chromosome 13q31-q33-related biallelic marker, or g35030-related biallelic marker in said sequence; optionally, wherein said polymorphism or biallelic marker is selected from the group consisting of A1 to A127, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said biallelic marker is selected from the group consisting of A13 to A18, A20 to A46, A49 to A52, A55, A57, A59 to A63, A72 to A73, A76 and A123, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said biallelic marker is selected from the group consisting of A13 to A65; optionally, wherein the 3' end of said polynucleotide is located 1 nucleotide upstream of said polymorphism or biallelic marker; and optionally, wherein said polynucleotide comprises, consists of, or consists essentially of a sequence selected from the following sequences: D1 to D127 and E1 to E127.

In a further embodiment, the invention encompasses isolated, purified, or recombinant polynucleotides comprising, consisting of, or consisting essentially of a sequence selected from the following sequences: B1 to B114 and C1 to C114.

In an additional embodiment, the invention encompasses polynucleotides for use in hybridization assays, sequencing assays, and enzyme-based mismatch detection assays for determining the identity of the nucleotide at a polymorphism or biallelic marker in any of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 or the complement thereof, as well as polynucleotides for use in amplifying segments of nucleotides comprising a polymorphism or biallelic marker of the invention in any of SEQ ID Nos 1 to 17, 24, 25, 28, 74 and 79 to 132 or the complement thereof; optionally, wherein said polymorphism or biallelic marker is selected from the group consisting of A1 to A127, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said biallelic marker

is selected from the group consisting of A13 to A18, A20 to A46, A49 to A52, A55, A57, A59 to A63, A72 to A73, A76 and A123, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith; optionally, wherein said biallelic marker is selected from the group consisting of A13 to A65, and the complements thereof, or optionally the biallelic markers in linkage disequilibrium therewith.

These arrays may generally be produced using mechanical synthesis methods or light directed synthesis methods, which incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis (Fodor et al., Science, 251:767-777, 1991). The immobilization of arrays of oligonucleotides on solid supports has been rendered possible by the development of a technology generally identified as "Very Large Scale Immobilized Polymer Synthesis" (VLSIPS™) in which, typically, probes are immobilized in a high density array on a solid surface of a chip. Examples of VLSIPS™ technologies are provided in US Patents 5,143,854 and 5,412,087 and in PCT Publications WO 90/15070, WO 92/10092 and WO 95/11995, which describe methods for forming oligonucleotide arrays through techniques such as light-directed synthesis technique. In designing strategies aimed at providing arrays of nucleotides immobilized on solid supports, further presentation strategies were developed to order and display the oligonucleotide arrays on the chips in an attempt to maximize hybridization patterns and sequence information. Examples of such presentation strategies are disclosed in PCT Publications WO 94/12305, WO 94/11530, WO 97/29212 and WO 97/31256.

Oligonucleotide arrays may comprise at least one of the sequences selected from the group consisting of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132; and the sequences complementary thereto or a fragment thereof of at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500, 1000 or 2000 consecutive nucleotides, to the extent that fragments of these lengths is consistent with the lengths of the particular Sequence ID, for determining whether a sample contains one or more alleles of the biallelic markers of the present invention.

Oligonucleotide arrays may also comprise at least one of the sequences selected from the group consisting of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132; and the sequences complementary thereto or a fragment thereof of at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500, 1000 or 2000 consecutive nucleotides, to the extent that fragments of these lengths is consistent with the lengths of the particular Sequence ID, for amplifying one or more alleles of the biallelic markers of Table 5. In other embodiments, arrays may also comprise at least one of the sequences selected from the group consisting of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132; and the sequences complementary thereto or a fragment thereof of at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500, 1000 or 2000 consecutive nucleotides, to the extent that fragments of these lengths is consistent with the lengths of the

particular Sequence ID, for conducting microsequencing analyses to determine whether a sample contains one or more alleles of the biallelic markers of the invention. In still further embodiments, the oligonucleotide array may comprise at least one of the sequences selecting from the group consisting of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132; and the sequences complementary thereto or a fragment thereof of at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500, 1000 or 2000 nucleotides in length, to the extent that fragments of these lengths is consistent with the lengths of the particular Sequence ID, for determining whether a sample contains one or more alleles of the polymorphisms and biallelic markers of the present invention.

A further object of the invention relates to an array of nucleic acid sequences comprising either at least one of the sequences selected from the group consisting of P1 to P127, B1 to B114, C1 to C114, D1 to D127 E1 to E127 or the sequences complementary thereto or a fragment thereof of at least 8, 10, 12, 15, 18, 20, 25, 30, or 40 consecutive nucleotides thereof, or at least one sequence comprising at least 1, 2, 3, 4, 5, 10, 20 biallelic markers selected from the group consisting of A1 to A127 or the complements thereof. The invention also pertains to an array of nucleic acid sequences comprising either at least 1, 2, 3, 4, 5, 10, 20 of the sequences selected from the group consisting of P1 to P127, B1 to B114, C1 to C114, D1 to D127, E1 to E127 or the sequences complementary thereto or a fragment thereof of at least 8 consecutive nucleotides thereof, or at least two sequences comprising a biallelic marker selected from the group consisting of A1 to A127 or the complements thereto. In preferred embodiments, said biallelic markers are selected from the group consisting of A13 to A18, A20 to A46, A49 to A52, A55, A57, A59 to A63, A72 to A73, A76 and A123. In preferred embodiments, said probes, amplification primers and microsequencing primers correspond in Table 6 to said biallelic markers.

The present invention also encompasses diagnostic kits comprising one or more polynucleotides of the invention, optionally with a portion or all of the necessary reagents and instructions for genotyping a test subject by determining the identity of a nucleotide at an biallelic marker of the invention. The polynucleotides of a kit may optionally be attached to a solid support, or be part of an array or addressable array of polynucleotides. The kit may provide for the determination of the identity of the nucleotide at a marker position by any method known in the art including, but not limited to, a sequencing assay method, a microsequencing assay method, a hybridization assay method, or enzyme-based mismatch detection assay. Optionally such a kit may include instructions for scoring the results of the determination with respect to the test subjects' predisposition to schizophrenia, or likely response to an agent acting on schizophrenia, or chances of suffering from side effects to an

agent acting on schizophrenia.

Optionally, in any embodiments described herein, a biallelic marker of the invention may comprise:

- 5 (a) a biallelic marker selected from the group consisting of g35030-related markers A13 to A65; (b) a biallelic marker selected from the group consisting of chromosome 13q31-q33-related-biallelic markers A13 to A18, A20 to A46, A49 to A52, A55, A57, A59 to A63, A72 to A73, A76 and A123; or (c) a biallelic marker selected from the group consisting of chromosome 13q31-q33-related-biallelic markers A13 to A18, A20 to A47, A49 to A52, A55, A57, A59 to A63. Furthermore, any of the embodiments of the invention may specifically exclude one or  
10 more of the biallelic markers selected from biallelic markers A1 to A127.

#### **Methods For *De Novo* Identification Of Biallelic Markers**

Any of a variety of methods can be used to screen a genomic fragment for single nucleotide polymorphisms such as differential hybridization with oligonucleotide probes,  
15 detection of changes in the mobility measured by gel electrophoresis or direct sequencing of the amplified nucleic acid. A preferred method for identifying biallelic markers involves comparative sequencing of genomic DNA fragments from an appropriate number of unrelated individuals.

In a first embodiment, DNA samples from unrelated individuals are pooled together,  
20 following which the genomic DNA of interest is amplified and sequenced. The nucleotide sequences thus obtained are then analyzed to identify significant polymorphisms. One of the major advantages of this method resides in the fact that the pooling of the DNA samples substantially reduces the number of DNA amplification reactions and sequencing reactions, which must be carried out. Moreover, this method is sufficiently sensitive so that a biallelic  
25 marker obtained thereby usually demonstrates a sufficient frequency of its less common allele to be useful in conducting association studies. Usually, the frequency of the least common allele of a biallelic marker identified by this method is at least 10%.

In a second embodiment, the DNA samples are not pooled and are therefore amplified and sequenced individually. This method is usually preferred when biallelic markers need to be  
30 identified in order to perform association studies within candidate genes. Preferably, highly relevant gene regions such as promoter regions or exon regions may be screened for biallelic markers. A biallelic marker obtained using this method may show a lower degree of informativeness for conducting association studies, e.g. if the frequency of its less frequent allele may be less than about 10%. Such a biallelic marker will however be sufficiently  
35 informative to conduct association studies and it will further be appreciated that including less

informative biallelic markers in the genetic analysis studies of the present invention, may allow in some cases the direct identification of causal mutations, which may, depending on their penetrance, be rare mutations.

5 The following is a description of the various parameters of a preferred method used by the inventors for the identification of the biallelic markers of the present invention.

#### Genomic DNA samples

The genomic DNA samples from which the biallelic markers of the present invention are generated are preferably obtained from unrelated individuals corresponding to a heterogeneous population of known ethnic background. The number of individuals from whom DNA samples are obtained can vary substantially, preferably from about 10 to about 1000, more preferably from about 50 to about 200 individuals. Usually, DNA samples are collected from at least about 100 individuals in order to have sufficient polymorphic diversity in a given population to identify as many markers as possible and to generate statistically significant results.

15 As for the source of the genomic DNA to be subjected to analysis, any test sample can be foreseen without any particular limitation. These test samples include biological samples, which can be tested by the methods of the present invention described herein, and include human and animal body fluids such as whole blood, serum, plasma, cerebrospinal fluid, urine, lymph fluids, and various external secretions of the respiratory, intestinal and genitourinary tracts, tears, saliva, milk, white blood cells, myelomas and the like; biological fluids such as cell culture supernatants; fixed tissue specimens including tumor and non-tumor tissue and lymph node tissues; bone marrow aspirates and fixed cell specimens. The preferred source of genomic DNA used in the present invention is from peripheral venous blood of each donor. Techniques to prepare genomic DNA from biological samples are well known to the skilled technician.

20

25 Details of a preferred embodiment are provided in Example 1. The person skilled in the art can choose to amplify pooled or unpooled DNA samples.

#### DNA Amplification

The identification of biallelic markers in a sample of genomic DNA may be facilitated through the use of DNA amplification methods. DNA samples can be pooled or unpooled for the amplification step. DNA amplification techniques are well known to those skilled in the art. Various methods to amplify DNA fragments carrying biallelic markers are further described hereinafter herein. The PCR technology is the preferred amplification technique used to identify new biallelic markers.

30

35 In a first embodiment, biallelic markers are identified using genomic sequence information generated by the inventors. Genomic DNA fragments, such as the inserts of the



BAC clones described above, are sequenced and used to design primers for the amplification of 500 bp fragments. These 500 bp fragments are amplified from genomic DNA and are scanned for biallelic markers. Primers may be designed using the OSP software (Hillier L. and Green P., 1991). All primers may contain, upstream of the specific target bases, a common oligonucleotide tail that serves as a sequencing primer. Those skilled in the art are familiar with primer extensions, which can be used for these purposes.

In another embodiment of the invention, genomic sequences of candidate genes are available in public databases allowing direct screening for biallelic markers. Preferred primers, useful for the amplification of genomic sequences encoding the candidate genes, focus on promoters, exons and splice sites of the genes. A biallelic marker present in these functional regions of the gene have a higher probability to be a causal mutation.

#### **Sequencing Of Amplified Genomic DNA And Identification Of Single Nucleotide Polymorphisms**

The amplification products generated as described above, are then sequenced using any method known and available to the skilled technician. Methods for sequencing DNA using either the dideoxy-mediated method (Sanger method) or the Maxam-Gilbert method are widely known to those of ordinary skill in the art. Such methods are for example disclosed in Maniatis et al. (Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Second Edition, 1989). Alternative approaches include hybridization to high-density DNA probe arrays as described in Chee et al. (Science 274, 610, 1996).

Preferably, the amplified DNA is subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol. The products of the sequencing reactions are run on sequencing gels and the sequences are determined using gel image analysis. The polymorphism search is based on the presence of superimposed peaks in the electrophoresis pattern resulting from different bases occurring at the same position. Because each dideoxy terminator is labeled with a different fluorescent molecule, the two peaks corresponding to a biallelic site present distinct colors corresponding to two different nucleotides at the same position on the sequence. However, the presence of two peaks can be an artifact due to background noise. To exclude such an artifact, the two DNA strands are sequenced and a comparison between the peaks is carried out. In order to be registered as a polymorphic sequence, the polymorphism has to be detected on both strands.

The above procedure permits those amplification products, which contain biallelic markers to be identified. The detection limit for the frequency of biallelic polymorphisms detected by sequencing pools of 100 individuals is approximately 0.1 for the minor allele, as verified by sequencing pools of known allelic frequencies. However, more than 90% of the

biallelic polymorphisms detected by the pooling method have a frequency for the minor allele higher than 0.25. Therefore, the biallelic markers selected by this method have a frequency of at least 0.1 for the minor allele and less than 0.9 for the major allele. Preferably at least 0.2 for the minor allele and less than 0.8 for the major allele, more preferably at least 0.3 for the minor allele and less than 0.7 for the major allele, thus a heterozygosity rate higher than 0.18, preferably higher than 0.32, more preferably higher than 0.42.

In another embodiment, biallelic markers are detected by sequencing individual DNA samples, the frequency of the minor allele of such a biallelic marker may be less than 0.1.

#### 10                    **Validation of the biallelic markers of the present invention**

The polymorphisms are evaluated for their usefulness as genetic markers by validating that both alleles are present in a population. Validation of the biallelic markers is accomplished by genotyping a group of individuals by a method of the invention and demonstrating that both alleles are present. Microsequencing is a preferred method of genotyping alleles. The validation by genotyping step may be performed on individual samples derived from each individual in the group or by genotyping a pooled sample derived from more than one individual. The group can be as small as one individual if that individual is heterozygous for the allele in question. Preferably the group contains at least three individuals, more preferably the group contains five or six individuals, so that a single validation test will be more likely to result in the validation of more of the biallelic markers that are being tested. It should be noted, however, that when the validation test is performed on a small group it may result in a false negative result if as a result of sampling error none of the individuals tested carries one of the two alleles. Thus, the validation process is less useful in demonstrating that a particular initial result is an artifact, than it is at demonstrating that there is a *bona fide* biallelic marker at a particular position in a sequence. All of the genotyping, haplotyping, association, and interaction study methods of the invention may optionally be performed solely with validated biallelic markers.

#### 25                    **Evaluation of the frequency of the biallelic markers of the present invention**

30                    The validated biallelic markers are further evaluated for their usefulness as genetic markers by determining the frequency of the least common allele at the biallelic marker site. The determination of the least common allele is accomplished by genotyping a group of individuals by a method of the invention and demonstrating that both alleles are present. This determination of frequency by genotyping step may be performed on individual samples derived from each individual in the group or by genotyping a pooled sample derived from more than

one individual. The group must be large enough to be representative of the population as a whole. Preferably the group contains at least 20 individuals, more preferably the group contains at least 50 individuals, most preferably the group contains at least 100 individuals. Of course the larger the group the greater the accuracy of the frequency determination because of reduced sampling error. A biallelic marker wherein the frequency of the less common allele is 30% or more is termed a "high quality biallelic marker." All of the genotyping, haplotyping, association, and interaction study methods of the invention may optionally be performed solely with high quality biallelic markers.

Another embodiment of the invention comprises methods of estimating the frequency of an allele in a population comprising genotyping individuals from said population for a 13q31-q33-related biallelic marker and determining the proportional representation of said biallelic marker in said population. In addition, the methods of estimating the frequency of an allele in a population encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination: Optionally, said 13q31-q33-related biallelic marker may be in a sequence selected individually or in any combination from the group consisting of SEQ Nos 1 to 17, 24, 25, 28, 74 and 79 to 132; and the complements thereof; optionally, said 13q31-q33-related biallelic marker may be selected from the biallelic markers described in Table 5; optionally, determining the frequency of a biallelic marker allele in a population may be accomplished by determining the identity of the nucleotides for both copies of said biallelic marker present in the genome of each individual in said population and calculating the proportional representation of said nucleotide at said 13q31-q33-related biallelic marker for the population; optionally, determining the frequency of a biallelic marker allele in a population may be accomplished by performing a genotyping method on a pooled biological sample derived from a representative number of individuals, or each individual, in said population, and calculating the proportional amount of said nucleotide compared with the total.

#### **Methods Of Genotyping An Individual For Biallelic Markers**

Methods are provided to genotype a biological sample for one or more biallelic markers of the present invention, all of which may be performed *in vitro*. Such methods of genotyping comprise determining the identity of a nucleotide at an biallelic marker of the invention by any method known in the art. These methods find use in genotyping case-control populations in association studies as well as individuals in the context of detection of alleles of biallelic markers which, are known to be associated with a given trait, in which case both copies of the biallelic marker present in individual's genome are determined so that an individual may be classified as homozygous or heterozygous for a particular allele.

These genotyping methods can be performed nucleic acid samples derived from a single individual or pooled DNA samples.

Genotyping can be performed using similar methods as those described above for the identification of the biallelic markers, or using other genotyping methods such as those further described below. In preferred embodiments, the comparison of sequences of amplified genomic fragments from different individuals is used to identify new biallelic markers whereas microsequencing is used for genotyping known biallelic markers in diagnostic and association study applications.

Another embodiment of the invention encompasses methods of genotyping a biological sample comprising determining the identity of a nucleotide at a 13q31-q33-related biallelic marker. Thus, the invention encompasses a method of genotyping an individual comprising: (a) obtaining a biological sample comprising a nucleic acid from said individual; (b) determining the identity of a polymorphic base at a biallelic marker from said nucleic acid; wherein said polymorphic base is selected from the group consisting of the first or second respective allele at the nucleotide position of the polymorphic base of SEQ ID No 1, 24, 25, 28, or 74 as indicated in table 5; wherein the identity of the polymorphic base determines the genotype of the individual at said position.

In addition, the genotyping methods of the invention encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination: Optionally, said 13q31-q33-related biallelic marker may be in a sequence selected individually or in any combination from the group consisting of SEQ ID Nos 1 to 17, 24, 25, 28, 74 and 79 to 132, and the complements thereof; optionally, said 13q31-q33-related biallelic marker may be selected individually or in any combination from the biallelic markers described in Table 5; optionally, said method further comprises determining the identity of a second nucleotide at said biallelic marker, wherein said first nucleotide and second nucleotide are not base paired (by Watson & Crick base pairing) to one another; optionally, said biological sample is derived from a single individual or subject; optionally, said method is performed *in vitro*; optionally, said biallelic marker is determined for both copies of said biallelic marker present in said individual's genome; optionally, said biological sample is derived from multiple subjects or individuals; optionally, said method further comprises amplifying a portion of said sequence comprising the biallelic marker prior to said determining step; optionally, wherein said amplifying is performed by PCR, LCR, or replication of a recombinant vector comprising an origin of replication and said portion in a host cell; optionally, wherein said determining is performed by a hybridization assay, sequencing assay, microsequencing assay, or an enzyme-based mismatch detection assay.

### Source of DNA for genotyping

Any source of nucleic acids, in purified or non-purified form, can be utilized as the starting nucleic acid, provided it contains or is suspected of containing the specific nucleic acid sequence desired. DNA or RNA may be extracted from cells, tissues, body fluids and the like as described herein. While nucleic acids for use in the genotyping methods of the invention can be derived from any mammalian source, the test subjects and individuals from which nucleic acid samples are taken are generally understood to be human.

### Amplification Of DNA Fragments Comprising Biallelic Markers

Methods and polynucleotides are provided to amplify a segment of nucleotides comprising one or more biallelic marker of the present invention. It will be appreciated that amplification of DNA fragments comprising biallelic markers may be used in various methods and for various purposes and is not restricted to genotyping. Nevertheless, many genotyping methods, although not all, require the previous amplification of the DNA region carrying the biallelic marker of interest. Such methods specifically increase the concentration or total number of sequences that span the biallelic marker or include that site and sequences located either distal or proximal to it. Diagnostic assays may also rely on amplification of DNA segments carrying a biallelic marker of the present invention.

Amplification of DNA may be achieved by any method known in the art. The established PCR (polymerase chain reaction) method or by developments thereof or alternatives. Amplification methods which can be utilized herein include but are not limited to Ligase Chain Reaction (LCR) as described in EP A 320 308 and EP A 439 182, Gap LCR (Wolcott, M.J.), the so-called "NASBA" or "3SR" technique described in Guatelli J.C. et al. (1990) and in Compton J. (1991), Q-beta amplification as described in EP A 4544 610, strand displacement amplification as described in Walker et al. (1996) and EP A 684 315 and, target mediated amplification as described in PCT Publication WO 9322461.

LCR and Gap LCR are exponential amplification techniques, both depend on DNA ligase to join adjacent primers annealed to a DNA molecule. In Ligase Chain Reaction (LCR), probe pairs are used which include two primary (first and second) and two secondary (third and fourth) probes, all of which are employed in molar excess to target. The first probe hybridizes to a first segment of the target strand and the second probe hybridizes to a second segment of the target strand, the first and second segments being contiguous so that the primary probes abut one another in 5' phosphate-3'hydroxyl relationship, and so that a ligase can covalently fuse or ligate the two probes into a fused product. In addition, a third (secondary) probe can hybridize to a portion of the first probe and a fourth (secondary) probe can hybridize to a portion of the second probe in a similar abutting fashion. Of course, if the target is initially double stranded,

the secondary probes also will hybridize to the target complement in the first instance. Once the ligated strand of primary probes is separated from the target strand, it will hybridize with the third and fourth probes which can be ligated to form a complementary, secondary ligated product. It is important to realize that the ligated products are functionally equivalent to either  
5 the target or its complement. By repeated cycles of hybridization and ligation, amplification of the target sequence is achieved. A method for multiplex LCR has also been described (WO 9320227). Gap LCR (GLCR) is a version of LCR where the probes are not adjacent but are separated by 2 to 3 bases.

For amplification of mRNAs, it is within the scope of the present invention to reverse  
10 transcribe mRNA into cDNA followed by polymerase chain reaction (RT-PCR); or, to use a single enzyme for both steps as described in U.S. Patent No. 5,322,770 or, to use Asymmetric Gap LCR (RT-AGLCR) as described by Marshall R.L. et al. (1994). AGLCR is a modification of GLCR that allows the amplification of RNA.

Some of these amplification methods are particularly suited for the detection of single  
15 nucleotide polymorphisms and allow the simultaneous amplification of a target sequence and the identification of the polymorphic nucleotide as it is further described herein.

The PCR technology is the preferred amplification technique used in the present invention. A variety of PCR techniques are familiar to those skilled in the art. For a review of PCR technology, see Molecular Cloning to Genetic Engineering White, B.A. Ed. (1997) and the  
20 publication entitled "PCR Methods and Applications" (1991, Cold Spring Harbor Laboratory Press). In each of these PCR procedures, PCR primers on either side of the nucleic acid sequences to be amplified are added to a suitably prepared nucleic acid sample along with dNTPs and a thermostable polymerase such as Taq polymerase, Pfu polymerase, or Vent polymerase. The nucleic acid in the sample is denatured and the PCR primers are specifically  
25 hybridized to complementary nucleic acid sequences in the sample. The hybridized primers are extended. Thereafter, another cycle of denaturation, hybridization, and extension is initiated. The cycles are repeated multiple times to produce an amplified fragment containing the nucleic acid sequence between the primer sites. PCR has further been described in several patents including US Patents 4,683,195, 4,683,202 and 4,965,188.

30 Primers can be prepared by any suitable method. As for example, direct chemical synthesis by a method such as the phosphodiester method of Narang S.A. et al. (1979), the phosphodiester method of Brown E.L. et al. (1979), the diethylphosphoramidite method of Beaucage et al. (1981) and the solid support method described in EP 0 707 592.

In some embodiments the present invention provides primers for amplifying a DNA  
35 fragment containing one or more biallelic markers of the present invention. It will be

appreciated that the primers listed are merely exemplary and that any other set of primers which produce amplification products containing one or more biallelic markers of the present invention.

5 The spacing of the primers determines the length of the segment to be amplified. In the context of the present invention amplified segments carrying biallelic markers can range in size from at least about 25 bp to 35 kbp. Amplification fragments from 25-3000 bp are typical, fragments from 50-1000 bp are preferred and fragments from 100-600 bp are highly preferred. It will be appreciated that amplification primers for the biallelic markers may be any sequence which allow the specific amplification of any DNA fragment carrying the markers.  
10 Amplification primers may be labeled or immobilized on a solid support as described in the section titled "Oligonucleotide Probes and Primers".

#### **Methods of Genotyping DNA samples for Biallelic Markers**

Any method known in the art can be used to identify the nucleotide present at a biallelic marker site. Since the biallelic marker allele to be detected has been identified and specified in  
15 the present invention, detection will prove simple for one of ordinary skill in the art by employing any of a number of techniques. Many genotyping methods require the previous amplification of the DNA region carrying the biallelic marker of interest. While the amplification of target or signal is often preferred at present, ultrasensitive detection methods which do not require amplification are also encompassed by the present genotyping methods.  
20 Methods well-known to those skilled in the art that can be used to detect biallelic polymorphisms include methods such as, conventional dot blot analyzes, single strand conformational polymorphism analysis (SSCP) described by Orita et al. (1989), denaturing gradient gel electrophoresis (DGGE), heteroduplex analysis, mismatch cleavage detection, and other conventional techniques as described in Sheffield, V.C. et al. (1991), White et al. (1992),  
25 Grompe, M. et al. (1989) and Grompe, M. (1993). Another method for determining the identity of the nucleotide present at a particular polymorphic site employs a specialized exonuclease-resistant nucleotide derivative as described in US patent 4,656,127.

Preferred methods involve directly determining the identity of the nucleotide present at a biallelic marker site by sequencing assay, enzyme-based mismatch detection assay, or  
30 hybridization assay. The following is a description of some preferred methods. A highly preferred method is the microsequencing technique. The term "sequencing assay" is used herein to refer to polymerase extension of duplex primer/template complexes and includes both traditional sequencing and microsequencing.

#### **1) Sequencing assays**

35 The nucleotide present at a polymorphic site can be determined by sequencing methods.

In a preferred embodiment, DNA samples are subjected to PCR amplification before sequencing as described above. DNA sequencing methods are described in herein. Preferably, the amplified DNA is subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol. Sequence analysis allows the identification of the base present at the biallelic marker site.

## 2) Microsequencing assays

In microsequencing methods, a nucleotide at the polymorphic site that is unique to one of the alleles in a target DNA is detected by a single nucleotide primer extension reaction. This method involves appropriate microsequencing primers which, hybridize just upstream of a polymorphic base of interest in the target nucleic acid. A polymerase is used to specifically extend the 3' end of the primer with one single ddNTP (chain terminator) complementary to the selected nucleotide at the polymorphic site. Next the identity of the incorporated nucleotide is determined in any suitable way.

Typically, microsequencing reactions are carried out using fluorescent ddNTPs and the extended microsequencing primers are analyzed by electrophoresis on ABI 377 sequencing machines to determine the identity of the incorporated nucleotide as described in EP 412 883. Alternatively capillary electrophoresis can be used in order to process a higher number of assays simultaneously. An example of a typical microsequencing procedure that can be used in the context of the present invention is provided in example 4.

Different approaches can be used to detect the nucleotide added to the microsequencing primer. A homogeneous phase detection method based on fluorescence resonance energy transfer has been described by Chen and Kwok (1997) and Chen et al. (1997). In this method amplified genomic DNA fragments containing polymorphic sites are incubated with a 5'-fluorescein-labeled primer in the presence of allelic dye-labeled dideoxyribonucleoside triphosphates and a modified Taq polymerase. The dye-labeled primer is extended one base by the dye-terminator specific for the allele present on the template. At the end of the genotyping reaction, the fluorescence intensities of the two dyes in the reaction mixture are analyzed directly without separation or purification. All these steps can be performed in the same tube and the fluorescence changes can be monitored in real time. Alternatively, the extended primer may be analyzed by MALDI-TOF Mass Spectrometry. The base at the polymorphic site is identified by the mass added onto the microsequencing primer (see Haff L.A. and Smirnov I.P., 1997).

Microsequencing may be achieved by the established microsequencing method or by developments or derivatives thereof. Alternative methods include several solid-phase microsequencing techniques. The basic microsequencing protocol is the same as described



previously, except that the method is conducted as a heterogenous phase assay, in which the primer or the target molecule is immobilized or captured onto a solid support. To simplify the primer separation and the terminal nucleotide addition analysis, oligonucleotides are attached to solid supports or are modified in such ways that permit affinity separation as well as polymerase extension. The 5' ends and internal nucleotides of synthetic oligonucleotides can be modified in a number of different ways to permit different affinity separation approaches, e.g., biotinylation. If a single affinity group is used on the oligonucleotides, the oligonucleotides can be separated from the incorporated terminator reagent. This eliminates the need of physical or size separation. More than one oligonucleotide can be separated from the terminator reagent and analyzed simultaneously if more than one affinity group is used. This permits the analysis of several nucleic acid species or more nucleic acid sequence information per extension reaction. The affinity group need not be on the priming oligonucleotide but could alternatively be present on the template. For example, immobilization can be carried out via an interaction between biotinylated DNA and streptavidin-coated microtitration wells or avidin-coated polystyrene particles. In the same manner oligonucleotides or templates may be attached to a solid support in a high-density format. In such solid phase microsequencing reactions, incorporated ddNTPs can be radiolabeled (Syvänen, 1994) or linked to fluorescein (Livak and Hainer, 1994). The detection of radiolabeled ddNTPs can be achieved through scintillation-based techniques. The detection of fluorescein-linked ddNTPs can be based on the binding of anti fluorescein antibody conjugated with alkaline phosphatase, followed by incubation with a chromogenic substrate (such as *p*-nitrophenyl phosphate). Other possible reporter-detection pairs include: ddNTP linked to dinitrophenyl (DNP) and anti-DNP alkaline phosphatase conjugate (Harju et al., 1993) or biotinylated ddNTP and horseradish peroxidase-conjugated streptavidin with *o*-phenylenediamine as a substrate (WO 92/15712). As yet another alternative solid-phase microsequencing procedure, Nyren et al. (1993) described a method relying on the detection of DNA polymerase activity by an enzymatic luminometric inorganic pyrophosphate detection assay (ELIDA).

Pastinen et al. (1997), describe a method for multiplex detection of single nucleotide polymorphism in which the solid phase minisequencing principle is applied to an oligonucleotide array format. High-density arrays of DNA probes attached to a solid support (DNA chips) are further described in herein.

In one aspect the present invention provides polynucleotides and methods to genotype one or more biallelic markers of the present invention by performing a microsequencing assay. Preferred microsequencing primers include those being featured Table 6. It will be appreciated that the microsequencing primers listed in Table 6 are merely exemplary and that, any primer

having a 3' end immediately adjacent to a polymorphic nucleotide may be used. Similarly, it will be appreciated that microsequencing analysis may be performed for any biallelic marker or any combination of biallelic markers of the present invention. One aspect of the present invention is a solid support which includes one or more microsequencing primers listed in Table 6, or fragments comprising at least 8, at least 12, at least 15, or at least 20 consecutive nucleotides thereof and having a 3' terminus immediately upstream of the corresponding biallelic marker, for determining the identity of a nucleotide at biallelic marker site.

### 3) Mismatch detection assays based on polymerases and ligases

In one aspect the present invention provides polynucleotides and methods to determine the allele of one or more biallelic markers of the present invention in a biological sample, by mismatch detection assays based on polymerases and/or ligases. These assays are based on the specificity of polymerases and ligases. Polymerization reactions places particularly stringent requirements on correct base pairing of the 3' end of the amplification primer and the joining of two oligonucleotides hybridized to a target DNA sequence is quite sensitive to mismatches close to the ligation site, especially at the 3' end. The terms "enzyme based mismatch detection assay" are used herein to refer to any method of determining the allele of a biallelic marker based on the specificity of ligases and polymerases. Preferred methods are described below. Methods, primers and various parameters to amplify DNA fragments comprising biallelic markers of the present invention are further described herein.

#### Allele specific amplification

Discrimination between the two alleles of a biallelic marker can also be achieved by allele specific amplification, a selective strategy, whereby one of the alleles is amplified without amplification of the other allele. This is accomplished by placing a polymorphic base at the 3' end of one of the amplification primers. Because the extension forms from the 3' end of the primer, a mismatch at or near this position has an inhibitory effect on amplification. Therefore, under appropriate amplification conditions, these primers only direct amplification on their complementary allele. Designing the appropriate allele-specific primer and the corresponding assay conditions are well with the ordinary skill in the art.

#### Ligation/amplification based methods

The "Oligonucleotide Ligation Assay" (OLA) uses two oligonucleotides which are designed to be capable of hybridizing to abutting sequences of a single strand of a target molecules. One of the oligonucleotides is biotinylated, and the other is detectably labeled. If the precise complementary sequence is found in a target molecule, the oligonucleotides will hybridize such that their termini abut, and create a ligation substrate that can be captured and detected. OLA is capable of detecting biallelic markers and may be advantageously combined

with PCR as described by Nickerson D.A. et al. (1990). In this method, PCR is used to achieve the exponential amplification of target DNA, which is then detected using OLA.

Other methods which are particularly suited for the detection of biallelic markers include LCR (ligase chain reaction), Gap LCR (GLCR) which are described herein. As mentioned above LCR uses two pairs of probes to exponentially amplify a specific target. The sequences of each pair of oligonucleotides, is selected to permit the pair to hybridize to abutting sequences of the same strand of the target. Such hybridization forms a substrate for a template-dependant ligase. In accordance with the present invention, LCR can be performed with oligonucleotides having the proximal and distal sequences of the same strand of a biallelic marker site. In one embodiment, either oligonucleotide will be designed to include the biallelic marker site. In such an embodiment, the reaction conditions are selected such that the oligonucleotides can be ligated together only if the target molecule either contains or lacks the specific nucleotide(s) that is complementary to the biallelic marker on the oligonucleotide. In an alternative embodiment, the oligonucleotides will not include the biallelic marker, such that when they hybridize to the target molecule, a "gap" is created as described in WO 90/01069. This gap is then "filled" with complementary dNTPs (as mediated by DNA polymerase), or by an additional pair of oligonucleotides. Thus at the end of each cycle, each single strand has a complement capable of serving as a target during the next cycle and exponential allele-specific amplification of the desired sequence is obtained.

Ligase/Polymerase-mediated Genetic Bit Analysis<sup>TM</sup> is another method for determining the identity of a nucleotide at a preselected site in a nucleic acid molecule (WO 95/21271). This method involves the incorporation of a nucleoside triphosphate that is complementary to the nucleotide present at the preselected site onto the terminus of a primer molecule, and their subsequent ligation to a second oligonucleotide. The reaction is monitored by detecting a specific label attached to the reaction's solid phase or by detection in solution.

#### 4) Hybridization assay methods

A preferred method of determining the identity of the nucleotide present at a biallelic marker site involves nucleic acid hybridization. The hybridization probes, which can be conveniently used in such reactions, preferably include the probes defined herein. Any hybridization assay may be used including Southern hybridization, Northern hybridization, dot blot hybridization and solid-phase hybridization (see Sambrook et al., Molecular Cloning - A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., 1989).

Hybridization refers to the formation of a duplex structure by two single stranded nucleic acids due to complementary base pairing. Hybridization can occur between exactly complementary nucleic acid strands or between nucleic acid strands that contain minor regions

of mismatch. Specific probes can be designed that hybridize to one form of a biallelic marker and not to the other and therefore are able to discriminate between different allelic forms. Allele-specific probes are often used in pairs, one member of a pair showing perfect match to a target sequence containing the original allele and the other showing a perfect match to the target sequence containing the alternative allele. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. Stringent, sequence specific hybridization conditions, under which a probe will hybridize only to the exactly complementary target sequence are well known in the art (Sambrook et al., Molecular Cloning – A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., 1989). Stringent conditions are sequence dependent and will be different in different circumstances. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T<sub>m</sub>) for the specific sequence at a defined ionic strength and pH. By way of example and not limitation, procedures using conditions of high stringency are as follows: Prehybridization of filters containing DNA is carried out for 8 h to overnight at 65°C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 µg/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65°C, the preferred hybridization temperature, in prehybridization mixture containing 100 µg/ml denatured salmon sperm DNA and 5-20 X 10<sup>6</sup> cpm of <sup>32</sup>P-labeled probe. Alternatively, the hybridization step can be performed at 65°C in the presence of SSC buffer, 1 x SSC corresponding to 0.15M NaCl and 0.05 M Na citrate. Subsequently, filter washes can be done at 37°C for 1 h in a solution containing 2X SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA, followed by a wash in 0.1X SSC at 50°C for 45 min. Alternatively, filter washes can be performed in a solution containing 2 x SSC and 0.1% SDS, or 0.5 x SSC and 0.1% SDS, or 0.1 x SSC and 0.1% SDS at 68°C for 15 minute intervals. Following the wash steps, the hybridized probes are detectable by autoradiography. By way of example and not limitation, procedures using conditions of intermediate stringency are as follows: Filters containing DNA are prehybridized, and then hybridized at a temperature of 60°C in the presence of a 5 x SSC buffer and labeled probe. Subsequently, filters washes are performed in a solution containing 2x SSC at 50°C and the hybridized probes are detectable by autoradiography. Other conditions of high and intermediate stringency which may be used are well known in the art and as cited in Sambrook et al. (Molecular Cloning - A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., 1989) and Ausubel et al. (Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y., 1989).

Although such hybridizations can be performed in solution, it is preferred to employ a

solid-phase hybridization assay. The target DNA comprising a biallelic marker of the present invention may be amplified prior to the hybridization reaction. The presence of a specific allele in the sample is determined by detecting the presence or the absence of stable hybrid duplexes formed between the probe and the target DNA. The detection of hybrid duplexes can be carried out by a number of methods. Various detection assay formats are well known which utilize detectable labels bound to either the target or the probe to enable detection of the hybrid duplexes. Typically, hybridization duplexes are separated from unhybridized nucleic acids and the labels bound to the duplexes are then detected. Those skilled in the art will recognize that wash steps may be employed to wash away excess target DNA or probe. Standard heterogeneous assay formats are suitable for detecting the hybrids using the labels present on the primers and probes.

Two recently developed assays allow hybridization-based allele discrimination with no need for separations or washes (see Landegren U. et al., 1998). The TaqMan assay takes advantage of the 5' nuclease activity of Taq DNA polymerase to digest a DNA probe annealed specifically to the accumulating amplification product. TaqMan probes are labeled with a donor-acceptor dye pair that interacts via fluorescence energy transfer. Cleavage of the TaqMan probe by the advancing polymerase during amplification dissociates the donor dye from the quenching acceptor dye, greatly increasing the donor fluorescence. All reagents necessary to detect two allelic variants can be assembled at the beginning of the reaction and the results are monitored in real time (see Livak et al, 1995). In an alternative homogeneous hybridization-based procedure, molecular beacons are used for allele discriminations. Molecular beacons are hairpin-shaped oligonucleotide probes that report the presence of specific nucleic acids in homogeneous solutions. When they bind to their targets they undergo a conformational reorganization that restores the fluorescence of an internally quenched fluorophore (Tyagi et al., 1998).

By assaying the hybridization to an allele specific probe, one can detect the presence or absence of a biallelic marker allele in a given sample.

High-Throughput parallel hybridizations in array format are specifically encompassed within "hybridization assays" and are described below.

#### **Hybridization to addressable arrays of oligonucleotides**

Hybridization assays based on oligonucleotide arrays rely on the differences in hybridization stability of short oligonucleotides to perfectly matched and mismatched target sequence variants. Efficient access to polymorphism information is obtained through a basic structure comprising high-density arrays of oligonucleotide probes attached to a solid support (the chip) at selected positions. Each DNA chip can contain thousands to millions of individual

synthetic DNA probes arranged in a grid-like pattern and miniaturized to the size of a dime.

The chip technology has already been applied with success in numerous cases. For example, the screening of mutations has been undertaken in the BRCA1 gene, in *S. cerevisiae* mutant strains, and in the protease gene of HIV-1 virus (Hacia et al., 1996; Shoemaker et al., 1996; Kozal et al., 1996). Chips of various formats for use in detecting biallelic polymorphisms can be produced on a customized basis by Affymetrix (GeneChip™), Hyseq (HyChip and HyGnostics), and Protogene Laboratories.

In general, these methods employ arrays of oligonucleotide probes that are complementary to target nucleic acid sequence segments from an individual which, target sequences include a polymorphic marker. EP785280, describes a tiling strategy for the detection of single nucleotide polymorphisms. Briefly, arrays may generally be "tilled" for a large number of specific polymorphisms. By "tiling" is generally meant the synthesis of a defined set of oligonucleotide probes which is made up of a sequence complementary to the target sequence of interest, as well as preselected variations of that sequence, e.g., substitution of one or more given positions with one or more members of the basis set of monomers, i.e. nucleotides. Tiling strategies are further described in PCT application No. WO 95/11995. In a particular aspect, arrays are tiled for a number of specific, identified biallelic marker sequences. In particular the array is tiled to include a number of detection blocks, each detection block being specific for a specific biallelic marker or a set of biallelic markers. For example, a detection block may be tiled to include a number of probes, which span the sequence segment that includes a specific polymorphism. To ensure probes that are complementary to each allele, the probes are synthesized in pairs differing at the biallelic marker. In addition to the probes differing at the polymorphic base, monosubstituted probes are also generally tiled within the detection block. These monosubstituted probes have bases at and up to a certain number of bases in either direction from the polymorphism, substituted with the remaining nucleotides (selected from A, T, G, C and U). Typically the probes in a tiled detection block will include substitutions of the sequence positions up to and including those that are 5 bases away from the biallelic marker. The monosubstituted probes provide internal controls for the tiled array, to distinguish actual hybridization from artefactual cross-hybridization. Upon completion of hybridization with the target sequence and washing of the array, the array is scanned to determine the position on the array to which the target sequence hybridizes. The hybridization data from the scanned array is then analyzed to identify which allele or alleles of the biallelic marker are present in the sample. Hybridization and scanning may be carried out as described in PCT application No. WO 92/10092 and WO 95/11995 and US patent No. 5,424,186.

Thus, in some embodiments, the chips may comprise an array of nucleic acid sequences

of fragments of about 15 nucleotides in length. In further embodiments, the chip may comprise an array including at least one of the sequences selected from the group consisting of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 and the sequences complementary thereto, or a fragment thereof at least about 8 consecutive nucleotides, preferably 10, 15, 20, more preferably 25, 30, 40, 47, or 50 consecutive nucleotides. In some embodiments, the chip may comprise an array of at least 2, 3, 4, 5, 6, 7, 8 or more of these polynucleotides of the invention. Solid supports and polynucleotides of the present invention attached to solid supports are further described in the section titled "Oligonucleotide probes and Primers".

#### 5) Integrated Systems

Another technique, which may be used to analyze polymorphisms, includes multicomponent integrated systems, which miniaturize and compartmentalize processes such as PCR and capillary electrophoresis reactions in a single functional device. An example of such technique is disclosed in US patent 5,589,136, which describes the integration of PCR amplification and capillary electrophoresis in chips.

Integrated systems can be envisaged mainly when microfluidic systems are used. These systems comprise a pattern of microchannels designed onto a glass, silicon, quartz, or plastic wafer included on a microchip. The movements of the samples are controlled by electric, electroosmotic or hydrostatic forces applied across different areas of the microchip. For genotyping biallelic markers, the microfluidic system may integrate nucleic acid amplification, microsequencing, capillary electrophoresis and a detection method such as laser-induced fluorescence detection.

#### Methods Of Genetic Analysis Using The Biallelic Markers Of The Present Invention

Different methods are available for the genetic analysis of complex traits (see Lander and Schork, 1994). The search for disease-susceptibility genes is conducted using two main methods: the linkage approach in which evidence is sought for cosegregation between a locus and a putative trait locus using family studies, and the association approach in which evidence is sought for a statistically significant association between an allele and a trait or a trait causing allele (Khoury J. et al, 1993). In general, the biallelic markers of the present invention find use in any method known in the art to demonstrate a statistically significant correlation between a genotype and a phenotype. The biallelic markers may be used in parametric and non-parametric linkage analysis methods. Preferably, the biallelic markers of the present invention are used to identify genes associated with detectable traits using association studies, an approach which does not require the use of affected families and which permits the identification of genes

associated with complex and sporadic traits.

The genetic analysis using the biallelic markers of the present invention may be conducted on any scale. The whole set of biallelic markers of the present invention or any subset of biallelic markers of the present invention may be used. In some embodiments a subset of biallelic markers corresponding to one or several candidate genes of the present invention may be used. Alternatively, a subset of biallelic markers of the present invention localised on a specific chromosome segment may be used. Further, any set of genetic markers including a biallelic marker of the present invention may be used. As mentioned above, it should be noted that the biallelic markers of the present invention may be included in any complete or partial genetic map of the human genome. These different uses are specifically contemplated in the present invention and claims.

#### **Linkage analysis**

Linkage analysis is based upon establishing a correlation between the transmission of genetic markers and that of a specific trait throughout generations within a family. Thus, the aim of linkage analysis is to detect marker loci that show cosegregation with a trait of interest in pedigrees.

#### **Parametric methods**

When data are available from successive generations there is the opportunity to study the degree of linkage between pairs of loci. Estimates of the recombination fraction enable loci to be ordered and placed onto a genetic map. With loci that are genetic markers, a genetic map can be established, and then the strength of linkage between markers and traits can be calculated and used to indicate the relative positions of markers and genes affecting those traits (Weir, B.S., 1996). The classical method for linkage analysis is the logarithm of odds (lod) score method (see Morton N.E., 1955; Ott J, 1991). Calculation of lod scores requires specification of the mode of inheritance for the disease (parametric method). Generally, the length of the candidate region identified using linkage analysis is between 2 and 20Mb. Once a candidate region is identified as described above, analysis of recombinant individuals using additional markers allows further delineation of the candidate region. Linkage analysis studies have generally relied on the use of a maximum of 5,000 microsatellite markers, thus limiting the maximum theoretical attainable resolution of linkage analysis to about 600 kb on average.

Linkage analysis has been successfully applied to map simple genetic traits that show clear Mendelian inheritance patterns and which have a high penetrance (i.e., the ratio between the number of trait positive carriers of allele *a* and the total number of *a* carriers in the population). However, parametric linkage analysis suffers from a variety of drawbacks. First, it



is limited by its reliance on the choice of a genetic model suitable for each studied trait. Furthermore, as already mentioned, the resolution attainable using linkage analysis is limited, and complementary studies are required to refine the analysis of the typical 2Mb to 20Mb regions initially identified through linkage analysis. In addition, parametric linkage analysis approaches have proven difficult when applied to complex genetic traits, such as those due to the combined action of multiple genes and/or environmental factors. It is very difficult to model these factors adequately in a lod score analysis. In such cases, too large an effort and cost are needed to recruit the adequate number of affected families required for applying linkage analysis to these situations, as recently discussed by Risch, N. and Merikangas, K. (1996).

#### 10                   **Non-parametric methods**

The advantage of the so-called non-parametric methods for linkage analysis is that they do not require specification of the mode of inheritance for the disease, they tend to be more useful for the analysis of complex traits. In non-parametric methods, one tries to prove that the inheritance pattern of a chromosomal region is not consistent with random Mendelian segregation by showing that affected relatives inherit identical copies of the region more often than expected by chance. Affected relatives should show excess "allele sharing" even in the presence of incomplete penetrance and polygenic inheritance. In non-parametric linkage analysis the degree of agreement at a marker locus in two individuals can be measured either by the number of alleles identical by state (IBS) or by the number of alleles identical by descent (IBD). Affected sib pair analysis is a well-known special case and is the simplest form of these methods.

The biallelic markers of the present invention may be used in both parametric and non-parametric linkage analysis. Preferably biallelic markers may be used in non-parametric methods which allow the mapping of genes involved in complex traits. The biallelic markers of the present invention may be used in both IBD- and IBS- methods to map genes affecting a complex trait. In such studies, taking advantage of the high density of biallelic markers, several adjacent biallelic marker loci may be pooled to achieve the efficiency attained by multi-allelic markers (Zhao et al., 1998).

However, both parametric and non-parametric linkage analysis methods analyse affected relatives, they tend to be of limited value in the genetic analysis of drug responses or in the analysis of side effects to treatments. This type of analysis is impractical in such cases due to the lack of availability of familial cases. In fact, the likelihood of having more than one individual in a family being exposed to the same drug at the same time is extremely low.

#### **Population Association Studies**

35                   The present invention comprises methods for identifying one or several genes among a

set of candidate genes that are associated with a detectable trait using the biallelic markers of the present invention. In one embodiment the present invention comprises methods to detect an association between a biallelic marker allele or a biallelic marker haplotype and a trait. Further, the invention comprises methods to identify a trait causing allele in linkage disequilibrium with any biallelic marker allele of the present invention.

As described above, alternative approaches can be employed to perform association studies: genome-wide association studies, candidate region association studies and candidate gene association studies. The candidate region analysis clearly provides a short-cut approach to the identification of genes and gene polymorphisms related to a particular trait when some information concerning the biology of the trait is available. Further, the biallelic markers of the present invention may be incorporated in any map of genetic markers of the human genome in order to perform genome-wide association studies. Methods to generate a high-density map of biallelic markers has been described in US Provisional Patent application serial number 60/082,614. The biallelic markers of the present invention may further be incorporated in any map of a specific candidate region of the genome (a specific chromosome or a specific chromosomal segment for example).

As mentioned above, association studies may be conducted within the general population and are not limited to studies performed on related individuals in affected families. Association studies are extremely valuable as they permit the analysis of sporadic or multifactor traits. Moreover, association studies represent a powerful method for fine-scale mapping enabling much finer mapping of trait causing alleles than linkage studies. Studies based on pedigrees often only narrow the location of the trait causing allele. Association studies using the biallelic markers of the present invention can therefore be used to refine the location of a trait causing allele in a candidate region identified by Linkage Analysis methods. Biallelic markers of the present invention can be used to identify the involved gene; such uses are specifically contemplated in the present invention and claims.

**1) Determining the frequency of a biallelic marker allele or of a biallelic marker haplotype in a population**

Another embodiment of the present invention encompasses methods of estimating the frequency of a haplotype for a set of biallelic markers in a population, comprising the steps of: a) genotyping each individual in said population for at least one 13q31-q33-related biallelic marker, b) genotyping each individual in said population for a second biallelic marker by determining the identity of the nucleotides at said second biallelic marker for both copies of said second biallelic marker present in the genome; and c) applying a haplotype determination method to the identities of the nucleotides determined in steps a) and b) to obtain an estimate of

said frequency. In addition, the methods of estimating the frequency of a haplotype of the invention encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination: optionally said haplotype determination method is selected from the group consisting of asymmetric PCR amplification, double PCR  
5 amplification of specific alleles, the Clark method, or an expectation maximization algorithm; optionally, said second biallelic marker is a 13q31-q33-related biallelic marker in a sequence selected from the group consisting of SEQ ID Nos 1 to 17, 24, 25, 28, 74 and 79 to 132, and the complements thereof; optionally, said 13q31-q33-related biallelic marker may be selected individually or in any combination from the biallelic markers described in Tables 6b and 6c;  
10 optionally, the identity of the nucleotides at the biallelic markers in everyone of the sequences of SEQ ID Nos 1 to 17, 24, 25, 28, 74 and 79 to 132 is determined in steps a) and b).

**Association studies explore the relationships among frequencies for sets of alleles between loci.**

**Determining the frequency of an allele in a population**

15 Allelic frequencies of the biallelic markers in a population can be determined using one of the methods described above under the heading "Methods for genotyping an individual for biallelic markers", or any genotyping procedure suitable for this intended purpose. Genotyping pooled samples or individual samples can determine the frequency of a biallelic marker allele in a population. One way to reduce the number of genotypings required is to use pooled samples.  
20 A major obstacle in using pooled samples is in terms of accuracy and reproducibility for determining accurate DNA concentrations in setting up the pools. Genotyping individual samples provides higher sensitivity, reproducibility and accuracy and; is the preferred method used in the present invention. Preferably, each individual is genotyped separately and simple gene counting is applied to determine the frequency of an allele of a biallelic marker or of a  
25 genotype in a given population.

**Determining the frequency of a haplotype in a population**

The gametic phase of haplotypes is unknown when diploid individuals are heterozygous at more than one locus. Using genealogical information in families gametic phase can sometimes be inferred (Perlin et al., 1994). When no genealogical information is available  
30 different strategies may be used. One possibility is that the multiple-site heterozygous diploids can be eliminated from the analysis, keeping only the homozygotes and the single-site heterozygote individuals, but this approach might lead to a possible bias in the sample composition and the underestimation of low-frequency haplotypes. Another possibility is that single chromosomes can be studied independently, for example, by asymmetric PCR  
35 amplification (see Newton et al., 1989; Wu et al., 1989) or by isolation of single chromosome

by limit dilution followed by PCR amplification (see Ruano et al., 1990). Further, a sample may be haplotyped for sufficiently close biallelic markers by double PCR amplification of specific alleles (Sarkar, G. and Sommer S.S., 1991). These approaches are not entirely satisfying either because of their technical complexity, the additional cost they entail, their lack of generalisation at a large scale, or the possible biases they introduce. To overcome these difficulties, an algorithm to infer the phase of PCR-amplified DNA genotypes introduced by Clark A.G. (1990) may be used. Briefly, the principle is to start filling a preliminary list of haplotypes present in the sample by examining unambiguous individuals, that is, the complete homozygotes and the single-site heterozygotes. Then other individuals in the same sample are screened for the possible occurrence of previously recognised haplotypes. For each positive identification, the complementary haplotype is added to the list of recognised haplotypes, until the phase information for all individuals is either resolved or identified as unresolved. This method assigns a single haplotype to each multiheterozygous individual, whereas several haplotypes are possible when there are more than one heterozygous site. Alternatively, one can use methods estimating haplotype frequencies in a population without assigning haplotypes to each individual. Preferably, a method based on an expectation-maximization (EM) algorithm (Dempster et al., J. R. 1977) leading to maximum-likelihood estimates of haplotype frequencies under the assumption of Hardy-Weinberg proportions (random mating) is used (see Excoffier L. and Slatkin M., 1995). The EM algorithm is a generalised iterative maximum-likelihood approach to estimation that is useful when data are ambiguous and/or incomplete. The EM algorithm is used to resolve heterozygotes into haplotypes. Haplotype estimations are further described below under the heading "Statistical methods". Any other method known in the art to determine or to estimate the frequency of a haplotype in a population may also be used.

## 2) Linkage Disequilibrium analysis

Linkage disequilibrium is the non-random association of alleles at two or more loci and represents a powerful tool for mapping genes involved in disease traits (see Ajioka R.S. et al., 1997). Biallelic markers, because they are densely spaced in the human genome and can be genotyped in more numerous numbers than other types of genetic markers (such as RFLP or VNTR markers), are particularly useful in genetic analysis based on linkage disequilibrium. The biallelic markers of the present invention may be used in any linkage disequilibrium analysis method known in the art.

Briefly, when a disease mutation is first introduced into a population (by a new mutation or the immigration of a mutation carrier), it necessarily resides on a single chromosome and thus on a single "background" or "ancestral" haplotype of linked markers. Consequently, there is complete disequilibrium between these markers and the disease

mutation: one finds the disease mutation only in the presence of a specific set of marker alleles. Through subsequent generations recombinations occur between the disease mutation and these marker polymorphisms, and the disequilibrium gradually dissipates. The pace of this dissipation is a function of the recombination frequency, so the markers closest to the disease gene will manifest higher levels of disequilibrium than those that are further away. When not broken up by recombination, "ancestral" haplotypes and linkage disequilibrium between marker alleles at different loci can be tracked not only through pedigrees but also through populations. Linkage disequilibrium is usually seen as an association between one specific allele at one locus and another specific allele at a second locus.

The pattern or curve of disequilibrium between disease and marker loci is expected to exhibit a maximum that occurs at the disease locus. Consequently, the amount of linkage disequilibrium between a disease allele and closely linked genetic markers may yield valuable information regarding the location of the disease gene. For fine-scale mapping of a disease locus, it is useful to have some knowledge of the patterns of linkage disequilibrium that exist between markers in the studied region. As mentioned above the mapping resolution achieved through the analysis of linkage disequilibrium is much higher than that of linkage studies. The high density of biallelic markers combined with linkage disequilibrium analysis provides powerful tools for fine-scale mapping. Different methods to calculate linkage disequilibrium are described below under the heading "Statistical Methods".

### **3) Population-based case-control studies of trait-marker associations**

As mentioned above, the occurrence of pairs of specific alleles at different loci on the same chromosome is not random and the deviation from random is called linkage disequilibrium. Association studies focus on population frequencies and rely on the phenomenon of linkage disequilibrium. If a specific allele in a given gene is directly involved in causing a particular trait, its frequency will be statistically increased in an affected (trait positive) population, when compared to the frequency in a trait negative population or in a random control population. As a consequence of the existence of linkage disequilibrium, the frequency of all other alleles present in the haplotype carrying the trait-causing allele will also be increased in trait positive individuals compared to trait negative individuals or random controls. Therefore, association between the trait and any allele (specifically a biallelic marker allele) in linkage disequilibrium with the trait-causing allele will suffice to suggest the presence of a trait-related gene in that particular region. Case-control populations can be genotyped for biallelic markers to identify associations that narrowly locate a trait causing allele. As any marker in linkage disequilibrium with one given marker associated with a trait will be associated with the trait. Linkage disequilibrium allows the relative frequencies in case-control

populations of a limited number of genetic polymorphisms (specifically biallelic markers) to be analysed as an alternative to screening all possible functional polymorphisms in order to find trait-causing alleles. Association studies compare the frequency of marker alleles in unrelated case-control populations, and represent powerful tools for the dissection of complex traits.

5           **Case-control populations (inclusion criteria)**

Population-based association studies do not concern familial inheritance but compare the prevalence of a particular genetic marker, or a set of markers, in case-control populations. They are case-control studies based on comparison of unrelated case (affected or trait positive) individuals and unrelated control (unaffected or trait negative or random) individuals.

10       Preferably the control group is composed of unaffected or trait negative individuals. Further, the control group is ethnically matched to the case population. Moreover, the control group is preferably matched to the case-population for the main known confusion factor for the trait under study (for example age-matched for an age-dependent trait). Ideally, individuals in the two samples are paired in such a way that they are expected to differ only in their disease status.  
15       In the following "trait positive population", "case population" and "affected population" are used interchangeably.

          An important step in the dissection of complex traits using association studies is the choice of case-control populations (see Lander and Schork, 1994). A major step in the choice of case-control populations is the clinical definition of a given trait or phenotype. Any genetic  
20       trait may be analysed by the association method proposed here by carefully selecting the individuals to be included in the trait positive and trait negative phenotypic groups. Four criteria are often useful: clinical phenotype, age at onset, family history and severity. The selection procedure for continuous or quantitative traits (such as blood pressure for example) involves selecting individuals at opposite ends of the phenotype distribution of the trait under  
25       study, so as to include in these trait positive and trait negative populations individuals with non-overlapping phenotypes. Preferably, case-control populations comprise phenotypically homogeneous populations. Trait positive and trait negative populations comprise phenotypically uniform populations of individuals representing each between 1 and 98%, preferably between 1 and 80%, more preferably between 1 and 50%, and more preferably  
30       between 1 and 30%, most preferably between 1 and 20% of the total population under study, and selected among individuals exhibiting non-overlapping phenotypes. The clearer the difference between the two trait phenotypes, the greater the probability of detecting an association with biallelic markers. The selection of those drastically different but relatively uniform phenotypes enables efficient comparisons in association studies and the possible  
35       detection of marked differences at the genetic level, provided that the sample sizes of the

populations under study are significant enough.

In preferred embodiments, a first group of between 50 and 300 trait positive individuals, preferably about 100 individuals, are recruited according to their phenotypes. A similar number of trait negative individuals are included in such studies.

5 In the present invention, typical examples of inclusion criteria include affection by schizophrenia.

#### Association analysis

The general strategy to perform association studies using biallelic markers derived from a region carrying a candidate gene is to scan two groups of individuals (case-control  
10 populations) in order to measure and statistically compare the allele frequencies of the biallelic markers of the present invention in both groups.

If a statistically significant association with a trait is identified for at least one or more of the analysed biallelic markers, one can assume that: either the associated allele is directly responsible for causing the trait (the associated allele is the trait causing allele), or more likely  
15 the associated allele is in linkage disequilibrium with the trait causing allele. The specific characteristics of the associated allele with respect to the gene function usually gives further insight into the relationship between the associated allele and the trait (causal or in linkage disequilibrium). If the evidence indicates that the associated allele within the gene is most probably not the trait causing allele but is in linkage disequilibrium with the real trait causing  
20 allele, then the trait causing allele can be found by sequencing the vicinity of the associated marker.

Another embodiment of the present invention encompasses methods of detecting an association between a haplotype and a phenotype, comprising the steps of: a) estimating the frequency of at least one haplotype in a trait positive population according to a method of  
25 estimating the frequency of a haplotype of the invention; b) estimating the frequency of said haplotype in a control population according to the method of estimating the frequency of a haplotype of the invention; and c) determining whether a statistically significant association exists between said haplotype and said phenotype. In addition, the methods of detecting an association between a haplotype and a phenotype of the invention encompass methods with any  
30 further limitation described in this disclosure, or those following, specified alone or in any combination: Optionally, said 13q31-q33-related biallelic marker may be in a sequence selected individually or in any combination from the group consisting of SEQ ID Nos 1 to 17, 24, 25, 28, 74 and 79 to 132, and the complements thereof; optionally, said 13q31-q33-related biallelic marker may be selected individually or in any combination from the biallelic markers  
35 described in Tables 6b and 6c; optionally, said control population may be a trait negative

population, or a random population; optionally, said phenotype is a disease involving schizophrenia, a response to an agent acting on schizophrenia, or a side effects to an agent acting on schizophrenia.

#### **Haplotype analysis**

5 As described above, when a chromosome carrying a disease allele first appears in a population as a result of either mutation or migration, the mutant allele necessarily resides on a chromosome having a set of linked markers: the ancestral haplotype. This haplotype can be tracked through populations and its statistical association with a given trait can be analysed. Complementing single point (allelic) association studies with multi-point association studies  
10 also called haplotype studies increases the statistical power of association studies. Thus, a haplotype association study allows one to define the frequency and the type of the ancestral carrier haplotype. A haplotype analysis is important in that it increases the statistical power of an analysis involving individual markers.

In a first stage of a haplotype frequency analysis, the frequency of the possible  
15 haplotypes based on various combinations of the identified biallelic markers of the invention is determined. The haplotype frequency is then compared for distinct populations of trait positive and control individuals. The number of trait positive individuals, which should be, subjected to this analysis to obtain statistically significant results usually ranges between 30 and 300, with a preferred number of individuals ranging between 50 and 150. The same considerations apply to  
20 the number of unaffected individuals (or random control) used in the study. The results of this first analysis provide haplotype frequencies in case-control populations, for each evaluated haplotype frequency a p-value and an odd ratio are calculated. If a statistically significant association is found the relative risk for an individual carrying the given haplotype of being affected with the trait under study can be approximated.

#### **Interaction Analysis**

25 The biallelic markers of the present invention may also be used to identify patterns of biallelic markers associated with detectable traits resulting from polygenic interactions. The analysis of genetic interaction between alleles at unlinked loci requires individual genotyping using the techniques described herein. The analysis of allelic interaction among a selected set  
30 of biallelic markers with appropriate level of statistical significance can be considered as a haplotype analysis. Interaction analysis comprises stratifying the case-control populations with respect to a given haplotype for the first loci and performing a haplotype analysis with the second loci with each subpopulation.

Statistical methods used in association studies are further described herein.

#### **4) Testing for linkage in the presence of association**



The biallelic markers of the present invention may further be used in TDT (transmission/disequilibrium test). TDT tests for both linkage and association and is not affected by population stratification. TDT requires data for affected individuals and their parents or data from unaffected sibs instead of from parents (see Spielmann S. et al., 1993; Schaid D.J. et al., 1996, Spielmann S. and Ewens W.J., 1998). Such combined tests generally reduce the false – positive errors produced by separate analyses.

#### **Statistical methods**

In general, any method known in the art to test whether a trait and a genotype show a statistically significant correlation may be used.

##### **1) Methods in linkage analysis**

Statistical methods and computer programs useful for linkage analysis are well-known to those skilled in the art (see Terwilliger J.D. and Ott J., 1994; Ott J., 1991).

##### **2) Methods to estimate haplotype frequencies in a population**

As described above, when genotypes are scored, it is often not possible to distinguish heterozygotes so that haplotype frequencies cannot be easily inferred. When the gametic phase is not known, haplotype frequencies can be estimated from the multilocus genotypic data. Any method known to person skilled in the art can be used to estimate haplotype frequencies (see Lange K., 1997; Weir, B.S., 1996) Preferably, maximum-likelihood haplotype frequencies are computed using an Expectation- Maximization (EM) algorithm (see Dempster et al., 1977; Excoffier L. and Slatkin M., 1995). This procedure is an iterative process aiming at obtaining maximum-likelihood estimates of haplotype frequencies from multi-locus genotype data when the gametic phase is unknown. Haplotype estimations are usually performed by applying the EM algorithm using for example the EM-HAPLO program (Hawley M.E. et al., 1994) or the Arlequin program (Schneider et al., 1997). The EM algorithm is a generalised iterative maximum likelihood approach to estimation and is briefly described below.

In the following part of this text, phenotypes will refer to multi-locus genotypes with unknown phase. Genotypes will refer to known-phase multi-locus genotypes. Suppose a sample of N unrelated individuals typed for K markers. The data observed are the unknown-phase K-locus phenotypes that can be categorised in F different phenotypes. Suppose that we have H underlying possible haplotypes (in case of K biallelic markers,  $H=2^K$ ).

For phenotype j, suppose that  $c_j$  genotypes are possible. We thus have the following equation

$$P_j = \sum_{i=1}^{c_j} pr(genotype_i) = \sum_{i=1}^{c_j} pr(h_k, h_l) \quad \text{Equation 1}$$

where  $P_j$  is the probability of the phenotype  $j$ ,  $h_k$  and  $h_l$  are the two haplotypes constituent the genotype  $i$ . Under the Hardy-Weinberg equilibrium,  $pr(h_k h_l)$  becomes :

$$pr(h_k, h_l) = pr(h_k)^2 \text{ if } h_k = h_l, pr(h_k, h_l) = 2 pr(h_k) \cdot pr(h_l) \text{ if } h_k \neq h_l. \text{ Equation 2}$$

The successive steps of the E-M algorithm can be described as follows:

- 5 Starting with initial values of the of haplotypes frequencies, noted  $p_1^{(0)}, p_2^{(0)}, \dots, p_H^{(0)}$ , these initial values serve to estimate the genotype frequencies (Expectation step) and then estimate another set of haplotype frequencies (Maximisation step), noted  $p_1^{(1)}, p_2^{(1)}, \dots, p_H^{(1)}$ , these two steps are iterated until changes in the sets of haplotypes frequency are very small.

- 10 A stop criterion can be that the maximum difference between haplotype frequencies between two iterations is less than  $10^{-7}$ . These values can be adjusted according to the desired precision of estimations. In details, at a given iteration  $s$ , the Expectation step comprises calculating the genotypes frequencies by the following equation:

$$\begin{aligned} pr(genotype_i)^{(s)} &= pr(phenotype_j) \cdot pr(genotype_i | phenotype_j)^{(s)} \\ &= \frac{n_j}{N} \cdot \frac{pr(h_k, h_l)^{(s)}}{p_j^{(s)}} \end{aligned} \quad \text{Equation 3}$$

- 15 where genotype  $i$  occurs in phenotype  $j$ , and where  $h_k$  and  $h_l$  constitute genotype  $i$ . Each probability is derived according to eq.1, and eq.2 described above.

Then the Maximisation step simply estimates another set of haplotype frequencies given the genotypes frequencies. This approach is also known as gene-counting method (Smith, 1957).

$$p_i^{(s+1)} = \frac{1}{2} \sum_{j=1}^F \sum_{l=1}^{c_j} \delta_{il} \cdot pr(genotype_i)^{(s)} \quad \text{Equation 4}$$

- 20 Where  $\delta_{il}$  is an indicator variable which count the number of time haplotype  $l$  in genotype  $i$ . It takes the values of 0, 1 or 2.

To ensure that the estimation finally obtained is the maximum-likelihood estimation several values of departures are required. The estimations obtained are compared and if they are different the estimations leading to the best likelihood are kept.

- 25 **3) Methods to calculate linkage disequilibrium between markers**

A number of methods can be used to calculate linkage disequilibrium between any two genetic positions, in practice linkage disequilibrium is measured by applying a statistical association test to haplotype data taken from a population. Linkage disequilibrium between any pair of biallelic markers comprising at least one of the biallelic markers of the present invention

( $M_i, M_j$ ) having alleles ( $a_i/b_i$ ) at marker  $M_i$  and alleles ( $a_j/b_j$ ) at marker  $M_j$  can be calculated for every allele combination ( $a_i, a_j, a_i, b_j, b_i, a_j$  and  $b_i, b_j$ ), according to the Piazza formula :

$\Delta_{aiaj} = \sqrt{\theta_4} - \sqrt{(\theta_4 + \theta_3)(\theta_4 + \theta_2)}$ , where :

$\theta_4 = - -$  = frequency of genotypes not having allele  $a_i$  at  $M_i$  and not having allele  $a_j$  at  $M_j$

5  $\theta_3 = - +$  = frequency of genotypes not having allele  $a_i$  at  $M_i$  and having allele  $a_j$  at  $M_j$

$\theta_2 = + -$  = frequency of genotypes having allele  $a_i$  at  $M_i$  and not having allele  $a_j$  at  $M_j$

Linkage disequilibrium (LD) between pairs of biallelic markers ( $M_i, M_j$ ) can also be calculated for every allele combination ( $a_i, a_j, a_i, b_j, b_i, a_j$  and  $b_i, b_j$ ), according to the maximum-likelihood estimate (MLE) for delta (the composite genotypic disequilibrium coefficient), as described by Weir (Weir B.S., 1996). The MLE for the composite linkage disequilibrium is:

$$D_{aiaj} = (2n_1 + n_2 + n_3 + n_4/2)/N - 2(pr(a_i).pr(a_j))$$

where  $n_1 = \Sigma$  phenotype ( $a_i/a_i, a_j/a_j$ ),  $n_2 = \Sigma$  phenotype ( $a_i/a_i, a_j/b_j$ ),  $n_3 = \Sigma$  phenotype ( $a_i/b_i, a_j/a_j$ ),  $n_4 = \Sigma$  phenotype ( $a_i/b_i, a_j/b_j$ ) and  $N$  is the number of individuals in the sample. This formula allows linkage disequilibrium between alleles to be estimated when only genotype, and not haplotype, data are available.

Another means of calculating the linkage disequilibrium between markers is as follows. For a couple of biallelic markers,  $M_i (a_i/b_i)$  and  $M_j (a_j/b_j)$ , fitting the Hardy-Weinberg equilibrium, one can estimate the four possible haplotype frequencies in a given population according to the approach described above.

20 The estimation of gametic disequilibrium between  $a_i$  and  $a_j$  is simply:

$$D_{aiaj} = pr(haplotype(a_i, a_j)) - pr(a_i).pr(a_j).$$

Where  $pr(a_i)$  is the probability of allele  $a_i$  and  $pr(a_j)$  is the probability of allele  $a_j$  and where  $pr(haplotype(a_i, a_j))$  is estimated as in Equation 3 above.

For a couple of biallelic marker only one measure of disequilibrium is necessary to describe the association between  $M_i$  and  $M_j$ .

Then a normalised value of the above is calculated as follows:

$$D'_{aiaj} = D_{aiaj} / \max(-pr(a_i).pr(a_j), -pr(b_i).pr(b_j)) \text{ with } D_{aiaj} < 0$$

$$D'_{aiaj} = D_{aiaj} / \max(pr(b_i).pr(a_j), pr(a_i).pr(b_j)) \text{ with } D_{aiaj} > 0$$

The skilled person will readily appreciate that other LD calculation methods can be used without undue experimentation.

Linkage disequilibrium among a set of biallelic markers having an adequate heterozygosity rate can be determined by genotyping between 50 and 1000 unrelated individuals, preferably between 75 and 200, more preferably around 100.

#### 4) Testing for association

Methods for determining the statistical significance of a correlation between a phenotype and a genotype, in this case an allele at a biallelic marker or a haplotype made up of such alleles, may be determined by any statistical test known in the art and with any accepted threshold of statistical significance being required. The application of particular methods and thresholds of significance are well within the skill of the ordinary practitioner of the art.

Testing for association is performed by determining the frequency of a biallelic marker allele in case and control populations and comparing these frequencies with a statistical test to determine if there is a statistically significant difference in frequency which would indicate a correlation between the trait and the biallelic marker allele under study. Similarly, a haplotype analysis is performed by estimating the frequencies of all possible haplotypes for a given set of biallelic markers in case and control populations, and comparing these frequencies with a statistical test to determine if there is a statistically significant correlation between the haplotype and the phenotype (trait) under study. Any statistical tool useful to test for a statistically significant association between a genotype and a phenotype may be used. Preferably the statistical test employed is a chi-square test with one degree of freedom. A P-value is calculated (the P-value is the probability that a statistic as large or larger than the observed one would occur by chance).

#### **Statistical significance**

In preferred embodiments, significance for diagnosis purposes, either as a positive basis for further diagnostic tests or as a preliminary starting point for early preventive therapy, the p value related to a biallelic marker association is preferably about  $1 \times 10^{-2}$  or less, more preferably about  $1 \times 10^{-4}$  or less, for a single biallelic marker analysis and about  $1 \times 10^{-3}$  or less, still more preferably  $1 \times 10^{-6}$  or less and most preferably of about  $1 \times 10^{-8}$  or less, for a haplotype analysis involving several markers. These values are believed to be applicable to any association studies involving single or multiple marker combinations.

The skilled person can use the range of values set forth above as a starting point in order to carry out association studies with biallelic markers of the present invention. In doing so, significant associations between the biallelic markers of the present invention and diseases involving schizophrenia can be revealed and used for diagnosis and drug screening purposes.

#### **Phenotypic permutation**

In order to confirm the statistical significance of the first stage haplotype analysis described above, it might be suitable to perform further analyses in which genotyping data from case-control individuals are pooled and randomised with respect to the trait phenotype. Each individual genotyping data is randomly allocated to two groups, which contain the same number of individuals as the case-control populations used to compile the data obtained in the first

stage. A second stage haplotype analysis is preferably run on these artificial groups, preferably for the markers included in the haplotype of the first stage analysis showing the highest relative risk coefficient. This experiment is reiterated preferably at least between 100 and 10000 times. The repeated iterations allow the determination of the percentage of obtained haplotypes with a significant p-value level.

#### Assessment of statistical association

To address the problem of false positives similar analysis may be performed with the same case-control populations in random genomic regions. Results in random regions and the candidate region are compared as described in US Provisional Patent Application entitled "Methods, software and apparatus for identifying genomic regions harbouring a gene associated with a detectable trait".

#### 5) Evaluation of risk factors

The association between a risk factor (in genetic epidemiology the risk factor is the presence or the absence of a certain allele or haplotype at marker loci) and a disease is measured by the odds ratio (OR) and by the relative risk (RR). If  $P(R^+)$  is the probability of developing the disease for individuals with R and  $P(R^-)$  is the probability for individuals without the risk factor, then the relative risk is simply the ratio of the two probabilities, that is:

$$RR = P(R^+)/P(R^-)$$

In case-control studies, direct measures of the relative risk cannot be obtained because of the sampling design. However, the odds ratio allows a good approximation of the relative risk for low-incidence diseases and can be calculated:

$$OR = \left[ \frac{F^+}{1 - F^+} \right] / \left[ \frac{F^-}{1 - F^-} \right]$$

$F^+$  is the frequency of the exposure to the risk factor in cases and  $F^-$  is the frequency of the exposure to the risk factor in controls.  $F^+$  and  $F^-$  are calculated using the allelic or haplotype frequencies of the study and further depend on the underlying genetic model (dominant, recessive, additive...).

One can further estimate the attributable risk (AR) which describes the proportion of individuals in a population exhibiting a trait due to a given risk factor. This measure is important in quantitating the role of a specific factor in disease etiology and in terms of the public health impact of a risk factor. The public health relevance of this measure lies in estimating the proportion of cases of disease in the population that could be prevented if the exposure of interest were absent. AR is determined as follows:

$$AR = P_E (RR-1) / (P_E (RR-1)+1)$$

AR is the risk attributable to a biallelic marker allele or a biallelic marker haplotype.  $P_E$  is the frequency of exposure to an allele or a haplotype within the population at large; and RR is the relative risk which, is approximated with the odds ratio when the trait under study has a relatively low incidence in the general population.

5           AR is the risk attributable to a biallelic marker allele or a biallelic marker haplotype.  $P_E$  is the frequency of exposure to an allele or a haplotype within the population at large; and RR is the relative risk which, is approximated with the odds ratio when the trait under study has a relatively low incidence in the general population.

10           **Association of Biallelic Markers of the Invention with Schizophrenia**

In the context of the present invention, an association between chromosome 13q31-q33-related biallelic markers and schizophrenia and bipolar disorder were established. Several association studies using different populations and screening samples thereof, and with different sets of biallelic markers distributed on the chromosome 13q31-q33 region were carried out.

15           Further details concerning these association studies and the results are provided herein in Examples 5a to 5e.

          This information is extremely valuable. The knowledge of a potential genetic predisposition to schizophrenia, even if this predisposition is not absolute, might contribute in a very significant manner to treatment efficacy of schizophrenia and to the development of new  
20           therapeutic and diagnostic tools.

**Identification Of Biallelic Markers In Linkage Disequilibrium With The Biallelic Markers of the Invention**

          Once a first biallelic marker has been identified in a genomic region of interest, the  
25           practitioner of ordinary skill in the art, using the teachings of the present invention, can easily identify additional biallelic markers in linkage disequilibrium with this first marker. As mentioned before, any marker in linkage disequilibrium with a first marker associated with a trait will be associated with the trait. Therefore, once an association has been demonstrated between a given biallelic marker and a trait, the discovery of additional biallelic markers  
30           associated with this trait is of great interest in order to increase the density of biallelic markers in this particular region. The causal gene or mutation will be found in the vicinity of the marker or set of markers showing the highest correlation with the trait.

          Identification of additional markers in linkage disequilibrium with a given marker involves: (a) amplifying a genomic fragment comprising a first biallelic marker from a plurality  
35           of individuals; (b) identifying of second biallelic markers in the genomic region harboring said

first biallelic marker; (c) conducting a linkage disequilibrium analysis between said first biallelic marker and second biallelic markers; and (d) selecting said second biallelic markers as being in linkage disequilibrium with said first marker. Subcombinations comprising steps (b) and (c) are also contemplated.

5           Methods to identify biallelic markers and to conduct linkage disequilibrium analysis are described herein and can be carried out by the skilled person without undue experimentation. The present invention then also concerns biallelic markers and other polymorphisms which are in linkage disequilibrium with the specific biallelic markers of the invention and which are expected to present similar characteristics in terms of their respective association with a given  
10       trait. In a preferred embodiment, the invention concerns biallelic markers which are in linkage disequilibrium with the specific biallelic markers.

#### **Identification Of Functional Mutations**

15       Once a positive association is confirmed with a biallelic marker of the present invention, the associated candidate gene sequence can be scanned for mutations by comparing the sequences of a selected number of trait positive and trait negative individuals. In a preferred embodiment, functional regions such as exons and splice sites, promoters and other regulatory regions of the gene are scanned for mutations. Preferably, trait positive individuals carry the haplotype shown to be associated with the trait and trait negative individuals do not carry the  
20       haplotype or allele associated with the trait. The mutation detection procedure is essentially similar to that used for biallelic site identification.

25       The method used to detect such mutations generally comprises the following steps: (a) amplification of a region of the candidate DNA sequence comprising a biallelic marker or a group of biallelic markers associated with the trait from DNA samples of trait positive patients and trait negative controls; (b) sequencing of the amplified region; (c) comparison of DNA sequences from trait-positive patients and trait-negative controls; and (d) determination of mutations specific to trait-positive patients. Subcombinations which comprise steps (b) and (c) are specifically contemplated.

30       It is preferred that candidate polymorphisms be then verified by screening a larger population of cases and controls by means of any genotyping procedure such as those described herein, preferably using a microsequencing technique in an individual test format. Polymorphisms are considered as candidate mutations when present in cases and controls at frequencies compatible with the expected association results.

35       Candidate polymorphisms and mutations of the g35030 nucleic acid sequences suspected of being involved in a predisposition to schizophrenia can be confirmed by screening

a larger population of affected and unaffected individuals using any of the genotyping procedures described herein. Preferably the microsequencing technique is used. Such polymorphisms are considered as candidate "trait-causing" mutations when they exhibit a statistically significant correlation with the detectable phenotype.

5

#### **Biallelic Markers Of The Invention In Methods Of Genetic Diagnostics**

The biallelic markers and other polymorphisms of the present invention can also be used to develop diagnostics tests capable of identifying individuals who express a detectable trait as the result of a specific genotype or individuals whose genotype places them at risk of developing a detectable trait at a subsequent time. The trait analyzed using the present diagnostics may be any detectable trait, including predisposition to schizophrenia, age of onset of detectable symptoms, a beneficial response to or side effects related to treatment against schizophrenia. Such a diagnosis can be useful in the monitoring, prognosis and/or prophylactic or curative therapy for schizophrenia.

10

The diagnostic techniques of the present invention may employ a variety of methodologies to determine whether a test subject has a genotype associated with an increased risk of developing a detectable trait or whether the individual suffers from a detectable trait as a result of a particular mutation, including methods which enable the analysis of individual chromosomes for haplotyping, such as family studies, single sperm DNA analysis or somatic hybrids.

15

20

The diagnostic techniques concern the detection of specific alleles present within the human chromosome 13q31-q33 region and optionally within a g35030 nucleic acid sequence. More particularly, the invention concerns the detection of a nucleic acid comprising at least one of the nucleotide sequences of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 or a fragment thereof or a complementary sequence thereto including the polymorphic base.

25

These methods involve obtaining a nucleic acid sample from the individual and, determining, whether the nucleic acid sample contains at least one allele or at least one biallelic marker haplotype, indicative of a risk of developing the trait or indicative that the individual expresses the trait as a result of possessing a particular the human chromosome 13q31-q33- or g35030-related polymorphism or mutation (trait-causing allele).

30

Preferably, in such diagnostic methods, a nucleic acid sample is obtained from the individual and this sample is genotyped using methods described above in "Methods Of Genotyping DNA Samples For Biallelic markers." The diagnostics may be based on a single biallelic marker or a on group of biallelic markers.

35

In each of these methods, a nucleic acid sample is obtained from the test subject and the



biallelic marker pattern of one or more of the biallelic markers of the invention is determined.

In one embodiment, a PCR amplification is conducted on the nucleic acid sample to amplify regions in which polymorphisms associated with a detectable phenotype have been identified. The amplification products are sequenced to determine whether the individual  
5 possesses one or more human chromosome 13q31-q33 region- or g35030-related polymorphisms associated with a detectable phenotype. The primers used to generate amplification products may comprise the primers listed in Table 4. Alternatively, the nucleic acid sample is subjected to microsequencing reactions as described above to determine whether the individual possesses one or more human chromosome 13q31-q33 region-related  
10 polymorphisms associated with a detectable phenotype resulting from a mutation or a polymorphism in the human chromosome 13q31-q33 region or g35030-related biallelic marker. The primers used in the microsequencing reactions may include the primers listed in 6. In another embodiment, the nucleic acid sample is contacted with one or more allele specific oligonucleotide probes which, specifically hybridize to one or more human chromosome 13q31-  
15 q33 region or g35030-related alleles associated with a detectable phenotype. The probes used in the hybridization assay may include the probes listed in Table 5. In another embodiment, the nucleic acid sample is contacted with a second oligonucleotide capable of producing an amplification product when used with the allele specific oligonucleotide in an amplification reaction. The presence of an amplification product in the amplification reaction indicates that  
20 the individual possesses one or more human chromosome 13q31-q33 region or g35030-related alleles associated with a detectable phenotype.

In a preferred embodiment the identity of the nucleotide present at, at least one, biallelic marker selected from the group consisting of A13 to A18, A20 to A46, A49 to A52, A55, A57, A59 to A63, A72 to A73, A76 and A123 and the complements thereof, is determined and the  
25 detectable trait is schizophrenia. Diagnostic kits comprise any of the polynucleotides of the present invention.

These diagnostic methods are extremely valuable as they can, in certain circumstances, be used to initiate preventive treatments or to allow an individual carrying a significant haplotype to foresee warning signs such as minor symptoms.

30 Diagnostics, which analyze and predict response to a drug or side effects to a drug, may be used to determine whether an individual should be treated with a particular drug. For example, if the diagnostic indicates a likelihood that an individual will respond positively to treatment with a particular drug, the drug may be administered to the individual. Conversely, if the diagnostic indicates that an individual is likely to respond negatively to treatment with a  
35 particular drug, an alternative course of treatment may be prescribed. A negative response may

be defined as either the absence of an efficacious response or the presence of toxic side effects.

Clinical drug trials represent another application for the markers of the present invention. One or more markers indicative of response to an agent acting against schizophrenia or to side effects to an agent acting against schizophrenia may be identified using the methods described above. Thereafter, potential participants in clinical trials of such an agent may be screened to identify those individuals most likely to respond favorably to the drug and exclude those likely to experience side effects. In that way, the effectiveness of drug treatment may be measured in individuals who respond positively to the drug, without lowering the measurement as a result of the inclusion of individuals who are unlikely to respond positively in the study and without risking undesirable safety problems.

#### PREVENTION, DIAGNOSIS AND TREATMENT OF PSYCHIATRIC DISEASE

An aspect of the present invention relates to the preparation of a medicament for the treatment of psychiatric disease, in particular schizophrenia and bipolar disorder. The present invention embodies medicaments acting on g35030.

In preferred embodiments, medicaments of the invention act on g35030, either directly or indirectly, by acting on the g35030 pathways. For example, the medicaments may modulate, and more preferably decrease the level of g35030 activity which occurs in a cell or particular tissue, or increase or decrease the activity of the g35030 protein. In certain embodiments, the invention thus comprises use of a compound capable of increasing or decreasing g35030 expression or g35030 protein activity in the preparation or manufacture of a medicament. Preferably, said compound is used for the treatment of a psychiatric disease, preferably for the treatment of schizophrenia or bipolar disorder. Preferably, said compound acts directly by binding to g35030 of a g35030 receptor.

Such medicaments may also increase or decrease the activity of a compound analogous to g35030, a compound comprising an amino acid sequence having at least 25% homology to a sequence selected from the group consisting of SEQ ID NOs. 18 to 23, a compound comprising an amino acid sequence having at least 50% homology to a sequence selected from the group consisting of SEQ ID NOs. 18 to 23, and a compound comprising an amino acid sequence having at least 80% homology to a sequence selected from the group consisting of SEQ ID NOs. 18 to 23.

Medicaments which increase or decrease the activity of these compounds in an individual may be used to ameliorate or prevent symptoms in individuals suffering from or predisposed to a psychiatric disease, as discussed above in the section entitled "indications".

Alternatively, g35030 activity may be increased or decreased by the expression of the

genes encoding the identified g35030-modulating compounds using gene therapy. Examples of vectors and promoters suitable for use in gene therapy are described above. G35030 activity may also be increased or decreased by preparing an antibody which binds to a g35030 peptide, a g35030 receptor or a protein related thereto, as well as fragments of these proteins. Such antibodies may modulate the interaction between g35030 and a g35030 receptor or a protein related thereto. Antibodies and methods of obtaining them are further described herein.

As described above, the present invention provides cellular assays for identifying compounds for the treatment of psychiatric disease. The assays are based on detection of g35030 expression, measurement of g35030 protein activity, or based on the determination of other suitable schizophrenia, bipolar disorder or related psychiatric disease endpoints. Compounds for the treatment of psychiatric disease include derivative proteins or peptides which are capable of inhibiting the activity of a wild type g35030 protein, which may be identified by determining their ability to bind a wild type g35030 protein. Compounds also include antibodies, and small molecules and drugs which may be obtained using a variety of synthetic approaches familiar to those skilled in the art, including combinatorial chemistry based techniques.

The invention further encompasses said methods for the prevention, treatment, and diagnosis of disease using any of the g35030 nucleic acids or proteins of the invention in analogous methods.

#### **G35030 in Methods of Diagnosis or Detecting Predisposition**

Individuals affected by or predisposed to schizophrenia and bipolar disorder may express abnormal levels of g35030. Individuals having increased or decreased g35030 activity in their plasma, body fluids, or body tissues may be at risk of developing schizophrenia, bipolar disorder or a variety of potentially related psychiatric conditions. In one aspect of the present invention is a method for determining whether an individual is at risk of suffering from or is currently suffering from schizophrenia, bipolar disorder or other psychotic disorders, mood disorders, autism, substance dependence or alcoholism, mental retardation, or other psychiatric diseases including cognitive, anxiety, eating, impulse-control, and personality disorders, as defined with the Diagnosis and Statistical Manual of Mental Disorders fourth edition (DSM-IV) classification, comprising determining whether the individual has an abnormal level of g35030 activity in plasma, body fluids, or body tissues. The level of g35030 or analogous compounds in plasma, body fluids, or body tissues may be determined using a variety approaches. In particular, the level may be determined using ELISA, Western Blots, or protein electrophoresis.

#### **Biallelic Markers Of The Invention In Methods Of Genetic Diagnostics**

The biallelic markers and other polymorphisms of the present invention can also be used to develop diagnostics tests capable of identifying individuals who express a detectable

trait as the result of a specific genotype or individuals whose genotype places them at risk of developing a detectable trait at a subsequent time. The trait analyzed using the present diagnostics may be used to diagnose any detectable trait, including predisposition to schizophrenia or bipolar disorder, age of onset of detectable symptoms, a beneficial response to or side effects related to treatment against schizophrenia or bipolar disorder. Such a diagnosis can be useful in the monitoring, prognosis and/or prophylactic or curative therapy for schizophrenia or bipolar disorder.

The diagnostic techniques of the present invention may employ a variety of methodologies to determine whether a test subject has a genotype associated with an increased risk of developing a detectable trait or whether the individual suffers from a detectable trait as a result of a particular mutation, including methods which enable the analysis of individual chromosomes for haplotyping, such as family studies, single sperm DNA analysis or somatic hybrids.

The diagnostic techniques concern the detection of specific alleles present within the human chromosome 13q31-q33 region, and optionally within a g35030 nucleic acid sequence. More particularly, the invention concerns the detection of a nucleic acid comprising at least one of the nucleotide sequences of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 or a fragment thereof or a complementary sequence thereto including the polymorphic base.

These methods involve obtaining a nucleic acid sample from the individual and, determining, whether the nucleic acid sample contains at least one allele or at least one biallelic marker haplotype, indicative of a risk of developing the trait or indicative that the individual expresses the trait as a result of possessing a particular the human chromosome 13q31-q33 region-related polymorphism or mutation (trait-causing allele).

Preferably, in such diagnostic methods, a nucleic acid sample is obtained from the individual and this sample is genotyped using methods described above in "Methods Of Genotyping DNA Samples For Biallelic markers." The diagnostics may be based on a single biallelic marker or a on group of biallelic markers.

In each of these methods, a nucleic acid sample is obtained from the test subject and the biallelic marker pattern of one or more of a biallelic marker of the invention is determined.

In one embodiment, a PCR amplification is conducted on the nucleic acid sample to amplify regions in which polymorphisms associated with a detectable phenotype have been identified. The amplification products are sequenced to determine whether the individual possesses one or more human chromosome 13q31-q33 region, g35030-related polymorphisms associated with a detectable phenotype. The primers used to generate amplification products may comprise the primers listed in Table 4. Alternatively, the nucleic acid sample is subjected

to microsequencing reactions as described above to determine whether the individual possesses one or more human chromosome 13q31-q33 region, g35030-related polymorphisms associated with a detectable phenotype resulting from a mutation or a polymorphism in the human chromosome 13q31-q33 region. The primers used in the microsequencing reactions may include the primers listed in Table 6. In another embodiment, the nucleic acid sample is contacted with one or more allele specific oligonucleotide probes which, specifically hybridize to one or more human chromosome 13q31-q33 region, g35030-related alleles associated with a detectable phenotype. The probes used in the hybridization assay may include the probes listed in 6b. In another embodiment, the nucleic acid sample is contacted with a second oligonucleotide capable of producing an amplification product when used with the allele specific oligonucleotide in an amplification reaction. The presence of an amplification product in the amplification reaction indicates that the individual possesses one or more human chromosome 13q31-q33 region, g35030-related alleles associated with a detectable phenotype. In a preferred embodiment, the detectable trait is schizophrenia or bipolar disorder. Diagnostic kits comprise any of the polynucleotides of the present invention.

These diagnostic methods are extremely valuable as they can, in certain circumstances, be used to initiate preventive treatments or to allow an individual carrying a significant haplotype to foresee warning signs such as minor symptoms.

Diagnostics, which analyze and predict response to a drug or side effects to a drug, may be used to determine whether an individual should be treated with a particular drug. For example, if the diagnostic indicates a likelihood that an individual will respond positively to treatment with a particular drug, the drug may be administered to the individual. Conversely, if the diagnostic indicates that an individual is likely to respond negatively to treatment with a particular drug, an alternative course of treatment may be prescribed. A negative response may be defined as either the absence of an efficacious response or the presence of toxic side effects.

Clinical drug trials represent another application for the markers of the present invention. One or more markers indicative of response to an agent acting against schizophrenia or to side effects to an agent acting against schizophrenia may be identified using the methods described above. Thereafter, potential participants in clinical trials of such an agent may be screened to identify those individuals most likely to respond favorably to the drug and exclude those likely to experience side effects. In that way, the effectiveness of drug treatment may be measured in individuals who respond positively to the drug, without lowering the measurement as a result of the inclusion of individuals who are unlikely to respond positively in the study and without risking undesirable safety problems.

**Prevention And Treatment Of Disease Using Biallelic Markers**

In large part because of the risk of suicide, the detection of susceptibility to schizophrenia, bipolar disorder as well as other psychiatric disease in individuals is very important. Consequently, the invention concerns a method for the treatment of schizophrenia or bipolar disorder, or a related disorder comprising the following steps:

- 5 - selecting an individual whose DNA comprises alleles of a biallelic marker or of a group of biallelic markers of the human chromosome 13q31-q33 region, preferably g35030-related markers associated with schizophrenia or bipolar disorder;
- following up said individual for the appearance (and optionally the development) of the symptoms related to schizophrenia or bipolar disorder; and
- 10 - administering a treatment acting against schizophrenia or bipolar disorder or against symptoms thereof to said individual at an appropriate stage of the disease.

Another embodiment of the present invention comprises a method for the treatment of schizophrenia or bipolar disorder comprising the following steps:

- 15 - selecting an individual whose DNA comprises alleles of a biallelic marker or of a group of biallelic markers, of the human chromosome 13q31-q33 region, and more preferably g35030-related markers associated with schizophrenia or bipolar disorder;
- administering a preventive treatment of schizophrenia or bipolar disorder to said individual.

In a further embodiment, the present invention concerns a method for the treatment of schizophrenia or bipolar disorder comprising the following steps:

- 20 - selecting an individual whose DNA comprises alleles of a biallelic marker or of a group of biallelic markers of the human chromosome 13q31-q33, and more preferably g35030-related markers associated with schizophrenia or bipolar disorder;
- administering a preventive treatment of schizophrenia or bipolar disorder to said individual;
- 25 - following up said individual for the appearance and the development of schizophrenia or bipolar disorder symptoms; and optionally
- administering a treatment acting against schizophrenia or bipolar disorder or against symptoms thereof to said individual at the appropriate stage of the disease.

For use in the determination of the course of treatment of an individual suffering from disease, the present invention also concerns a method for the treatment of schizophrenia or bipolar disorder comprising the following steps:

- 30 - selecting an individual suffering from schizophrenia or bipolar disorder whose DNA comprises alleles of a biallelic marker or of a group of biallelic markers of the human chromosome 13q31-q33 region, and preferably g35030-related markers, associated with the
- 35 gravity of schizophrenia or bipolar disorder or of the symptoms thereof; and

- administering a treatment acting against schizophrenia or bipolar disorder or symptoms thereof to said individual.

The invention also concerns a method for the treatment of schizophrenia or bipolar disorder in a selected population of individuals. The method comprises:

- 5       - selecting an individual suffering from schizophrenia or bipolar disorder and whose DNA comprises alleles of a biallelic marker or of a group of biallelic markers of the human chromosome 13q31-q33 region, and more preferably g35030-related markers associated with a positive response to treatment with an effective amount of a medicament acting against schizophrenia or bipolar disorder or symptoms thereof,
- 10       - and/or whose DNA does not comprise alleles of a biallelic marker or of a group of biallelic markers of the human chromosome 13q31-q33 region, and more preferably g35030-related markers associated with a negative response to treatment with said medicament; and
- administering at suitable intervals an effective amount of said medicament to said selected individual.

15       In the context of the present invention, a "positive response" to a medicament can be defined as comprising a reduction of the symptoms related to the disease. In the context of the present invention, a "negative response" to a medicament can be defined as comprising either a lack of positive response to the medicament which does not lead to a symptom reduction or which leads to a side-effect observed following administration of the medicament.

20       The invention also relates to a method of determining whether a subject is likely to respond positively to treatment with a medicament. The method comprises identifying a first population of individuals who respond positively to said medicament and a second population of individuals who respond negatively to said medicament. One or more biallelic markers is identified in the first population which is associated with a positive response to said medicament

25       or one or more biallelic markers is identified in the second population which is associated with a negative response to said medicament. The biallelic markers may be identified using the techniques described herein.

      A DNA sample is then obtained from the subject to be tested. The DNA sample is analyzed to determine whether it comprises alleles of one or more biallelic markers associated

30       with a positive response to treatment with the medicament and/or alleles of one or more biallelic markers associated with a negative response to treatment with the medicament.

      In some embodiments, the medicament may be administered to the subject in a clinical trial if the DNA sample contains alleles of one or more biallelic markers associated with a positive response to treatment with the medicament and/or if the DNA sample lacks alleles of

35       one or more biallelic markers associated with a negative response to treatment with the

medicament. In preferred embodiments, the medicament is a drug acting against schizophrenia or bipolar disorder.

Using the method of the present invention, the evaluation of drug efficacy may be conducted in a population of individuals likely to respond favorably to the medicament.

5 Another aspect of the invention is a method of using a medicament comprising obtaining a DNA sample from a subject, determining whether the DNA sample contains alleles of one or more biallelic markers associated with a positive response to the medicament and/or whether the DNA sample contains alleles of one or more biallelic markers associated with a negative response to the medicament, and administering the medicament to the subject if the  
10 DNA sample contains alleles of one or more biallelic markers associated with a positive response to the medicament and/or if the DNA sample lacks alleles of one or more biallelic markers associated with a negative response to the medicament.

The invention also concerns a method for the clinical testing of a medicament, preferably a medicament acting against schizophrenia or or bipolar disorder or symptoms thereof. The method comprises the following steps:  
15

- administering a medicament, preferably a medicament susceptible of acting against schizophrenia or or bipolar disorder or symptoms thereof to a heterogeneous population of individuals,
- identifying a first population of individuals who respond positively to said  
20 medicament and a second population of individuals who respond negatively to said medicament,
- identifying biallelic markers in said first population which are associated with a positive response to said medicament,
- selecting individuals whose DNA comprises biallelic markers associated with a  
25 positive response to said medicament, and
- administering said medicament to said individuals.

In any of the methods for the prevention, diagnosis and treatment of schizophrenia and bipolar disorder, including methods of using a medicament, clinical testing of a medicament, determining whether a subject is likely to respond positively to treatment with a medicament, said biallelic marker may optionally comprise:  
30

- (a) a biallelic marker selected from the group consisting of g35030-related markers A13 to A65; (b) a biallelic marker selected from the group consisting of chromosome 13q31-q33-related-biallelic markers A13 to A18, A20 to A46, A49 to A52, A55, A57, A59 to A63, A72 to A73, A76 and A123 ; or (c) a biallelic marker selected from the group consisting of  
35 chromosome 13q31-q33-related-biallelic markers A13 to A18, A20 to A47, A49 to A52, A55,



A57, A59 to A63. Furthermore, any of the embodiments of the invention may specifically exclude one or more of the biallelic markers selected from biallelic markers A1 to A127.

Such methods are deemed to be extremely useful to increase the benefit/risk ratio resulting from the administration of medicaments which may cause undesirable side effects and/or be inefficacious to a portion of the patient population to which it is normally administered.

Once an individual has been diagnosed as suffering from schizophrenia or bipolar disorder, selection tests are carried out to determine whether the DNA of this individual comprises alleles of a biallelic marker or of a group of biallelic markers associated with a positive response to treatment or with a negative response to treatment which may include either side effects or unresponsiveness.

The selection of the patient to be treated using the method of the present invention can be carried out through the detection methods described above. The individuals which are to be selected are preferably those whose DNA does not comprise alleles of a biallelic marker or of a group of biallelic markers associated with a negative response to treatment. The knowledge of an individual's genetic predisposition to unresponsiveness or side effects to particular medicaments allows the clinician to direct treatment toward appropriate drugs against schizophrenia or bipolar disorder or symptoms thereof.

Once the patient's genetic predispositions have been determined, the clinician can select appropriate treatment for which negative response, particularly side effects, has not been reported or has been reported only marginally for the patient.

The biallelic markers of the invention have demonstrated an association with schizophrenia and bipolar disorders. However, the present invention also comprises any of the prevention, diagnostic, prognosis and treatment methods described herein using the biallelic markers of the invention in methods of preventing, diagnosing, managing and treating related disorders, particularly related CNS disorders. By way of example, related disorders may comprise psychotic disorders, mood disorders, autism, substance dependence and alcoholism, mental retardation, and other psychiatric diseases including cognitive, anxiety, eating, impulse-control, and personality disorders, as defined with the Diagnosis and Statistical Manual of Mental Disorders fourth edition (DSM-IV) classification".

#### **Recombinant Vectors**

The term "vector" is used herein to designate either a circular or a linear DNA or RNA molecule, which is either double-stranded or single-stranded, and which comprise at least one polynucleotide of interest that is sought to be transferred in a cell host or in a unicellular or multicellular host organism.

The present invention encompasses a family of recombinant vectors that comprise a polynucleotide derived from a g35030 nucleic acid sequence. Consequently, the present invention further comprises recombinant vectors comprising: g35030 genomic DNA or cDNAs comprised in the nucleic acids of any of nucleotide positions 201123 to 201234, 201123 to 201560, 214676 to 214793, 215702 to 215746, 216836 to 216994, 216836 to 217077, 217671 to 217764, 227655 to 227736, 238715 to 238919, 240440 to 240673, 240440 to 241153, 241072 to 241291, 244353 to 244561, 246273 to 247802, 201123 to 247802, 199122 to 201122, 247803 to 249803 and 199122 to 249803 of SEQ ID No. 1, SEQ ID Nos 2 to 17 and primate g35030 DNAs of SEQ ID Nos 79 to 132, and the complements thereof;

Generally, a recombinant vector of the invention may comprise any of the polynucleotides described herein, as well as any g35030 primer or probe as defined above.

In a first preferred embodiment, a recombinant vector of the invention is used to amplify the inserted polynucleotide derived from a g35030 genomic sequence or cDNA of the invention in a suitable cell host, this polynucleotide being amplified at every time that the recombinant vector replicates.

A second preferred embodiment of the recombinant vectors according to the invention comprises expression vectors comprising either a regulatory polynucleotide or a coding nucleic acid of the invention, or both. Within certain embodiments, expression vectors are employed to express a g35030 polypeptide which can be then purified and, for example be used in ligand screening assays or as an immunogen in order to raise specific antibodies directed against a g35030 protein. In other embodiments, the expression vectors are used for constructing transgenic animals and also for gene therapy. Expression requires that appropriate signals are provided in the vectors, said signals including various regulatory elements, such as enhancers/promoters from both viral and mammalian sources that drive expression of the genes of interest in host cells. Dominant drug selection markers for establishing permanent, stable cell clones expressing the products are generally included in the expression vectors of the invention, as they are elements that link expression of the drug selection markers to expression of the polypeptide.

More particularly, the present invention relates to expression vectors which include nucleic acids encoding a g35030 protein or variants or fragments thereof, under the control of a regulatory sequence of the respective g35030 regulatory polynucleotides, or alternatively under the control of an exogenous regulatory sequence.

The invention also pertains to a recombinant expression vector useful for the expression of a g35030 cDNA sequence.

Recombinant vectors comprising a nucleic acid containing a human chromosome

13q31-33-related biallelic marker, or more preferably a g35030-related biallelic marker is also part of the invention. In a preferred embodiment, said biallelic marker is selected from the group consisting of A13 to A65 and the complements thereof.

5 Some of the elements which can be found in the vectors of the present invention are described in further detail in the following sections.

### 1. General features of the expression vectors of the invention

A recombinant vector according to the invention comprises, but is not limited to, a YAC (Yeast Artificial Chromosome), a BAC (Bacterial Artificial Chromosome), a phage, a phagemid, a cosmid, a plasmid or even a linear DNA molecule which may comprise a  
10 chromosomal, non-chromosomal, semi-synthetic and synthetic DNA. Such a recombinant vector can comprise a transcriptional unit comprising an assembly of:

- (1) a genetic element or elements having a regulatory role in gene expression, for example promoters or enhancers. Enhancers are cis-acting elements of DNA, usually from about 10 to 300 bp in length that act on the promoter to increase the transcription.
- 15 (2) a structural or coding sequence which is transcribed into mRNA and eventually translated into a polypeptide, said structural or coding sequence being operably linked to the regulatory elements described in (1); and
- (3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence  
20 enabling extracellular secretion of translated protein by a host cell. Alternatively, when a recombinant protein is expressed without a leader or transport sequence, it may include a N-terminal residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

Generally, recombinant expression vectors will include origins of replication, selectable  
25 markers permitting transformation of the host cell, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably a leader sequence capable of directing secretion of the translated protein into the periplasmic space or the extracellular medium. In a specific embodiment  
30 wherein the vector is adapted for transfecting and expressing desired sequences in mammalian host cells, preferred vectors will comprise an origin of replication in the desired host, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5'-flanking non-transcribed sequences. DNA sequences derived from the SV40 viral genome, for example  
35 SV40 origin, early promoter, enhancer, splice and polyadenylation sites may be used to provide

the required non-transcribed genetic elements.

The *in vivo* expression of a g35030 polypeptide or fragments or variants thereof may be useful in order to correct a genetic defect related to the expression of the native gene in a host organism or to the production of a biologically inactive g35030 protein.

5           Consequently, the present invention also comprises recombinant expression vectors mainly designed for the *in vivo* production of the g35030 polypeptide by the introduction of the appropriate genetic material in the organism of the patient to be treated. In preferred embodiments, said genetic material comprises at least one nucleotide sequence selected from the group of nucleotide position ranges consisting of: g35030 genomic DNA or cDNAs  
10           comprised in the nucleic acids of any of nucleotide positions 201123 to 201234, 201123 to 201560, 214676 to 214793, 215702 to 215746, 216836 to 216994, 216836 to 217077, 217671 to 217764, 227655 to 227736, 238715 to 238919, 240440 to 240673, 240440 to 241153, 241072 to 241291, 244353 to 244561, 246273 to 247802, 201123 to 247802, 199122 to 201122, 247803 to 249803 and 199122 to 249803 of SEQ ID No. 1, SEQ ID Nos 2 to 17 and  
15           primate g35030 DNAs of SEQ ID Nos 79 to 132, and the complements thereof;

This genetic material may be introduced *in vitro* in a cell that has been previously extracted from the organism, the modified cell being subsequently reintroduced in the said organism, directly *in vivo* into the appropriate tissue.

## 2. Regulatory Elements

### 20           Promoters

The suitable promoter regions used in the expression vectors according to the present invention are chosen taking into account the cell host in which the heterologous gene has to be expressed. The particular promoter employed to control the expression of a nucleic acid sequence of interest is not believed to be important, so long as it is capable of directing the  
25           expression of the nucleic acid in the targeted cell. Thus, where a human cell is targeted, it is preferable to position the nucleic acid coding region adjacent to and under the control of a promoter that is capable of being expressed in a human cell, such as, for example, a human or a viral promoter.

A suitable promoter may be heterologous with respect to the nucleic acid for which it  
30           controls the expression or alternatively can be endogenous to the native polynucleotide containing the coding sequence to be expressed. Additionally, the promoter is generally heterologous with respect to the recombinant vector sequences within which the construct promoter/coding sequence has been inserted.

Promoter regions can be selected from any desired gene using, for example, CAT  
35           (chloramphenicol transferase) vectors and more preferably pKK232-8 and pCM7 vectors.

Preferred bacterial promoters are the LacI, LacZ, the T3 or T7 bacteriophage RNA polymerase promoters, the gpt, lambda PR, PL and trp promoters (EP 0036776), the polyhedrin promoter, or the p10 protein promoter from baculovirus (Kit Novagen) (Smith et al., 1983; O'Reilly et al., 1992), the lambda PR promoter or also the trc promoter.

5 Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-L. Selection of a convenient vector and promoter is well within the level of ordinary skill in the art.

The choice of a promoter is well within the ability of a person skilled in the field of genetic engineering. For example, one may refer to the book of Sambrook et al.(1989) or also  
10 to the procedures described by Fuller et al.(1996).

#### Other regulatory elements

One will typically desire to include a polyadenylation signal to effect proper polyadenylation of the gene transcript. The nature of the polyadenylation signal is not believed to be crucial to the successful practice of the invention, and any such sequence may be  
15 employed such as human growth hormone and SV40 polyadenylation signals. Also contemplated as an element of the expression cassette is a terminator. These elements can serve to enhance message levels and to minimize read through from the cassette into other sequences.

The vector containing the appropriate DNA sequence as described above, more preferably a g35030 gene regulatory polynucleotide, a polynucleotide encoding a g35030  
20 polypeptide comprising at least one nucleotide sequence selected from the group of nucleotide sequence ranges consisting of: g35030 genomic DNA or cDNAs comprised in the nucleic acids of any of nucleotide positions 201123 to 201234, 201123 to 201560, 214676 to 214793, 215702 to 215746, 216836 to 216994, 216836 to 217077, 217671 to 217764, 227655 to 227736, 238715 to 238919, 240440 to 240673, 240440 to 241153, 241072 to 241291, 244353 to  
25 244561, 246273 to 247802, 201123 to 247802, 199122 to 201122, 247803 to 249803 and 199122 to 249803 of SEQ ID No. 1, SEQ ID Nos 2 to 17 and primate g35030 DNAs of SEQ ID Nos 79 to 132, and the complements thereof.

#### **3. Selectable Markers**

Such markers would confer an identifiable change to the cell permitting easy  
30 identification of cells containing the expression construct. The selectable marker genes for selection of transformed host cells are preferably dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, TRP1 for *S. cerevisiae* or tetracycline, rifampicin or ampicillin resistance in *E. coli*, or levan saccharase for mycobacteria, this latter marker being a negative selection marker.

35 **4. Preferred Vectors.**

### **Bacterial vectors**

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and a bacterial origin of replication derived from commercially available plasmids comprising genetic elements of pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia, Uppsala, Sweden), and GEM1 (Promega Biotec, Madison, WI, USA).

Large numbers of other suitable vectors are known to those of skill in the art, and commercially available, such as the following bacterial vectors: pQE70, pQE60, pQE-9 (Qiagen), pbs, pD10, phagescript, psiX174, pbluescript SK, pbsks, pNH8A, pNH16A, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene); pSVK3, pBPV, pMSG, pSVL (Pharmacia); pQE-30 (QIAexpress).

### **Bacteriophage vectors**

The P1 bacteriophage vector may contain large inserts ranging from about 80 to about 100 kb.

The construction of P1 bacteriophage vectors such as p158 or p158/neo8 are notably described by Sternberg (1992, 1994). Recombinant P1 clones comprising g35030 polynucleotide sequences may be designed for inserting large polynucleotides of more than 40 kb (Linton et al., 1993). To generate P1 DNA for transgenic experiments, a preferred protocol is the protocol described by McCormick et al. (1994). Briefly, *E. coli* (preferably strain NS3529) harboring the P1 plasmid are grown overnight in a suitable broth medium containing 25 µg/ml of kanamycin. The P1 DNA is prepared from the *E. coli* by alkaline lysis using the Qiagen Plasmid Maxi kit (Qiagen, Chatsworth, CA, USA), according to the manufacturer's instructions. The P1 DNA is purified from the bacterial lysate on two Qiagen-tip 500 columns, using the washing and elution buffers contained in the kit. A phenol/chloroform extraction is then performed before precipitating the DNA with 70% ethanol. After solubilizing the DNA in TE (10 mM Tris-HCl, pH 7.4, 1 mM EDTA), the concentration of the DNA is assessed by spectrophotometry.

When the goal is to express a P1 clone comprising a g35030 polynucleotide sequence in a transgenic animal, typically in transgenic mice, it is desirable to remove vector sequences from the P1 DNA fragment, for example by cleaving the P1 DNA at rare-cutting sites within the P1 polylinker (*Sfi*I, *Not*I or *Sal*I). The P1 insert is then purified from vector sequences on a pulsed-field agarose gel, using methods similar to those originally reported for the isolation of DNA from YACs (Schedl et al., 1993a; Peterson et al., 1993, ). At this stage, the resulting purified insert DNA can be concentrated, if necessary, on a Millipore

Ultrafree-MC Filter Unit (Millipore, Bedford, MA, USA - 30,000 molecular weight limit) and then dialyzed against microinjection buffer (10 mM Tris-HCl, pH 7.4; 250  $\mu$ M EDTA) containing 100-mM NaCl, 30  $\mu$ M spermine, 70  $\mu$ M spermidine on a microdialysis membrane (type VS, 0.025  $\mu$ M from Millipore). The intactness of the purified P1 DNA insert is assessed  
5 by electrophoresis on 1% agarose (Sea Kem GTG; FMC Bio-products) pulse-field gel and staining with ethidium bromide.

#### Baculovirus vectors

A suitable vector for the expression of a g35030 polypeptide encoded by a polynucleotide of SEQ ID Nos. 1 to 17 or fragments or variants thereof is a baculovirus vector  
10 that can be propagated in insect cells and in insect cell lines. A specific suitable host vector system is the pVL1392/1393 baculovirus transfer vector (Pharminogen) that is used to transfect the SF9 cell line (ATCC N<sup>o</sup>CRL 1711) which is derived from *Spodoptera frugiperda*.

Other suitable vectors for the expression of the g35030 polypeptide encoded by the polynucleotides of SEQ ID Nos. 1 to 17, or fragments or variants thereof in a baculovirus  
15 expression system include those described by Chai et al.(1993), Vlasak et al.(1983) and Lenhard et al.(1996).

#### Viral vectors

In one specific embodiment, the vector is derived from an adenovirus. Preferred adenovirus vectors according to the invention are those described by Feldman and Steg (1996)  
20 or Ohno et al.(1994). Another preferred recombinant adenovirus according to this specific embodiment of the present invention is the human adenovirus type 2 or 5 (Ad 2 or Ad 5) or an adenovirus of animal origin (French patent application N<sup>o</sup> FR-93.05954).

Retrovirus vectors and adeno-associated virus vectors are generally understood to be the recombinant gene delivery systems of choice for the transfer of exogenous polynucleotides  
25 *in vivo*, particularly to mammals, including humans. These vectors provide efficient delivery of genes into cells, and the transferred nucleic acids are stably integrated into the chromosomal DNA of the host.

Particularly preferred retroviruses for the preparation or construction of retroviral *in vitro* or *in vitro* gene delivery vehicles of the present invention include retroviruses selected  
30 from the group consisting of Mink-Cell Focus Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis virus and Rous Sarcoma virus. Particularly preferred Murine Leukemia Viruses include the 4070A and the 1504A viruses, Abelson (ATCC No VR-999), Friend (ATCC No VR-245), Gross (ATCC No VR-590), Rauscher (ATCC No VR-998) and Moloney Murine Leukemia Virus (ATCC No VR-190; PCT Application No WO 94/24298). Particularly  
35 preferred Rous Sarcoma Viruses include Bryan high titer (ATCC Nos VR-334, VR-657, VR-

726, VR-659 and VR-728). Other preferred retroviral vectors are those described in Roth et al.(1996), PCT Application No WO 93/25234, PCT Application No WO 94/ 06920, Roux et al., 1989, Julian et al., 1992 and Neda et al., 1991.

5 Yet another viral vector system that is contemplated by the invention comprises the adeno-associated virus (AAV). The adeno-associated virus is a naturally occurring defective virus that requires another virus, such as an adenovirus or a herpes virus, as a helper virus for efficient replication and a productive life cycle (Muzyczka et al., 1992). It is also one of the few viruses that may integrate its DNA into non-dividing cells, and exhibits a high frequency of stable integration (Flotte et al., 1992; Samulski et al., 1989; McLaughlin et al., 1989). One  
10 advantageous feature of AAV derives from its reduced efficacy for transducing primary cells relative to transformed cells.

#### BAC vectors

The bacterial artificial chromosome (BAC) cloning system (Shizuya et al., 1992) has been developed to stably maintain large fragments of genomic DNA (100-300 kb) in *E. coli*. A  
15 preferred BAC vector comprises pBeloBAC11 vector that has been described by Kim et al.(1996). BAC libraries are prepared with this vector using size-selected genomic DNA that has been partially digested using enzymes that permit ligation into either the *Bam* HI or *Hind*III sites in the vector. Flanking these cloning sites are T7 and SP6 RNA polymerase transcription initiation sites that can be used to generate end probes by either RNA transcription or PCR  
20 methods. After the construction of a BAC library in *E. coli*, BAC DNA is purified from the host cell as a supercoiled circle. Converting these circular molecules into a linear form precedes both size determination and introduction of the BACs into recipient cells. The cloning site is flanked by two *Not* I sites, permitting cloned segments to be excised from the vector by *Not* I digestion. Alternatively, the DNA insert contained in the pBeloBAC11 vector may be  
25 linearized by treatment of the BAC vector with the commercially available enzyme lambda terminase that leads to the cleavage at the unique *cos*N site, but this cleavage method results in a full length BAC clone containing both the insert DNA and the BAC sequences.

#### **5. Delivery Of The Recombinant Vectors**

In order to effect expression of the polynucleotides and polynucleotide constructs of the  
30 invention, these constructs must be delivered into a cell. This delivery may be accomplished *in vitro*, as in laboratory procedures for transforming cell lines, or *in vivo* or *ex vivo*, as in the treatment of certain diseases states.

One mechanism is viral infection where the expression construct is encapsulated in an infectious viral particle.

35 Several non-viral methods for the transfer of polynucleotides into cultured mammalian



cells are also contemplated by the present invention, and include, without being limited to, calcium phosphate precipitation (Graham et al., 1973; Chen et al., 1987), DEAE-dextran (Gopal, 1985), electroporation (Tur-Kaspa et al., 1986; Potter et al., 1984), direct microinjection (Harland et al., 1985), DNA-loaded liposomes (Nicolau et al., 1982; Fraley et al., 1979), and receptor-mediate transfection (Wu and Wu, 1987; 1988). Some of these techniques may be successfully adapted for *in vivo* or *ex vivo* use.

Once the expression polynucleotide has been delivered into the cell, it may be stably integrated into the genome of the recipient cell. This integration may be in the cognate location and orientation via homologous recombination (gene replacement) or it may be integrated in a random, non specific location (gene augmentation). In yet further embodiments, the nucleic acid may be stably maintained in the cell as a separate, episomal segment of DNA. Such nucleic acid segments or "episomes" encode sequences sufficient to permit maintenance and replication independent of or in synchronization with the host cell cycle.

One specific embodiment for a method for delivering a protein or peptide to the interior of a cell of a vertebrate *in vivo* comprises the step of introducing a preparation comprising a physiologically acceptable carrier and a naked polynucleotide operatively coding for the polypeptide of interest into the interstitial space of a tissue comprising the cell, whereby the naked polynucleotide is taken up into the interior of the cell and has a physiological effect. This is particularly applicable for transfer *in vitro* but it may be applied to *in vivo* as well.

Compositions for use *in vitro* and *in vivo* comprising a "naked" polynucleotide are described in PCT application N° WO 90/11092 (Vical Inc.) and also in PCT application No. WO 95/11307 (Institut Pasteur, INSERM, Université d'Ottawa) as well as in the articles of Tacson et al.(1996) and of Huygen et al.(1996).

In still another embodiment of the invention, the transfer of a naked polynucleotide of the invention, including a polynucleotide construct of the invention, into cells may be proceeded with a particle bombardment (biolistic), said particles being DNA-coated microprojectiles accelerated to a high velocity allowing them to pierce cell membranes and enter cells without killing them, such as described by Klein et al.(1987).

In a further embodiment, the polynucleotide of the invention may be entrapped in a liposome (Ghosh and Bacchawat, 1991; Wong et al., 1980; Nicolau et al., 1987).

In a specific embodiment, the invention provides a composition for the *in vivo* production of the g35030 protein or polypeptide described herein. It comprises a naked polynucleotide operatively coding for this polypeptide, in solution in a physiologically acceptable carrier, and suitable for introduction into a tissue to cause cells of the tissue to express the said protein or polypeptide.

The amount of vector to be injected to the desired host organism varies according to the site of injection. As an indicative dose, it will be injected between 0,1 and 100 µg of the vector in an animal body, preferably a mammal body, for example a mouse body.

In another embodiment of the vector according to the invention, it may be introduced *in vitro* in a host cell, preferably in a host cell previously harvested from the animal to be treated and more preferably a somatic cell such as a muscle cell. In a subsequent step, the cell that has been transformed with the vector coding for the desired g35030 polypeptide or the desired fragment thereof is reintroduced into the animal body in order to deliver the recombinant protein within the body either locally or systemically.

#### Cell Hosts

Another object of the invention comprises a host cell that have been transformed or transfected with one of the polynucleotides described herein, and more precisely a polynucleotide comprising a g35030 polynucleotide selected from the group consisting of SEQ ID Nos. 1 to 17 and 79 to 132, or a fragment or a variant thereof. Are included host cells that are transformed (prokaryotic cells) or that are transfected (eukaryotic cells) with a recombinant vector such as one of those described above.

Generally, a recombinant host cell of the invention comprises any one of the polynucleotides or the recombinant vectors described therein.

Preferred host cells used as recipients for the expression vectors of the invention are the following:

a) Prokaryotic host cells: *Escherichia coli* strains (I.E.DH5-α strain), *Bacillus subtilis*, *Salmonella typhimurium*, and strains from species like *Pseudomonas*, *Streptomyces* and *Staphylococcus*.

b) Eukaryotic host cells: HeLa cells (ATCC N°CCL2; N°CCL2.1; N°CCL2.2), C<sub>v</sub> 1 cells (ATCC N°CCL70), COS cells (ATCC N°CRL1650; N°CRL1651), Sf-9 cells (ATCC N°CRL1711), C127 cells (ATCC N° CRL-1804), 3T3 (ATCC N° CRL-6361), CHO (ATCC N° CCL-61), human kidney 293. (ATCC N° 45504; N° CRL-1573) and BHK (ECACC N° 84100501; N° 84111301).

c) Other mammalian host cells.

G35030 gene expression in mammalian, and typically human, cells may be rendered defective with the replacement of a g35030 nucleic acid counterpart in the genome of an animal cell by a g35030 polynucleotide according to the invention. These genetic alterations may be generated by homologous recombination events using specific DNA constructs that have been previously described.

One kind of cell hosts that may be used are mammal zygotes, such as murine zygotes.

For example, murine zygotes may undergo microinjection with a purified DNA molecule of interest, for example a purified DNA molecule that has previously been adjusted to a concentration range from 1 ng/ml –for BAC inserts- 3 ng/ $\mu$ l –for P1 bacteriophage inserts- in 10 mM Tris-HCl, pH 7.4, 250  $\mu$ M EDTA containing 100 mM NaCl, 30  $\mu$ M spermine, and 70  $\mu$ M spermidine. When the DNA to be microinjected has a large size, polyamines and high salt concentrations can be used in order to avoid mechanical breakage of this DNA, as described by Schedl et al (1993b).

Any of the polynucleotides of the invention, including the DNA constructs described herein, may be introduced in an embryonic stem (ES) cell line, preferably a mouse ES cell line. ES cell lines are derived from pluripotent, uncommitted cells of the inner cell mass of pre-implantation blastocysts. Preferred ES cell lines are the following: ES-E14TG2a (ATCC n° CRL-1821), ES-D3 (ATCC n° CRL1934 and n° CRL-11632), YS001 (ATCC n° CRL-11776), 36.5 (ATCC n° CRL-11116). To maintain ES cells in an uncommitted state, they are cultured in the presence of growth inhibited feeder cells which provide the appropriate signals to preserve this embryonic phenotype and serve as a matrix for ES cell adherence. Preferred feeder cells are primary embryonic fibroblasts that are established from tissue of day 13- day 14 embryos of virtually any mouse strain, that are maintained in culture, such as described by Abbondanzo et al.(1993) and are inhibited in growth by irradiation, such as described by Robertson (1987), or by the presence of an inhibitory concentration of LIF, such as described by Pease and Williams (1990).

The constructs in the host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence.

Following transformation of a suitable host and growth of the host to an appropriate cell density, the selected promoter is induced by appropriate means, such as temperature shift or chemical induction, and cells are cultivated for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in the expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Such methods are well known by the skill artisan.

#### Transgenic Animals

The terms "transgenic animals" or "host animals" are used herein designate animals that have their genome genetically and artificially manipulated so as to include one of the nucleic acids according to the invention. Preferred animals are non-human mammals and include those belonging to a genus selected from *Mus* (e.g. mice), *Rattus* (e.g. rats) and *Oryctogalus* (e.g.

rabbits) which have their genome artificially and genetically altered by the insertion of a nucleic acid according to the invention. In one embodiment, the invention encompasses non-human host mammals and animals comprising a recombinant vector of the invention or a g35030 gene disrupted by homologous recombination with a knock out vector. The invention  
5 also encompasses non-human primates comprising a recombinant vector of the invention or a g35030 gene disrupted by homologous recombination with a knock out vector.

The transgenic animals of the invention all include within a plurality of their cells a cloned recombinant or synthetic DNA sequence, more specifically one of the purified or isolated nucleic acids comprising a g35030 polynucleotide or a DNA sequence encoding an  
10 antisense polynucleotide such as described in the present specification.

Generally, a transgenic animal according the present invention comprises any one of the polynucleotides, the recombinant vectors and the cell hosts described in the present invention.

In a first preferred embodiment, these transgenic animals may be good experimental models in order to study the diverse pathologies related to cell differentiation, in particular  
15 concerning the transgenic animals within the genome of which has been inserted one or several copies of a polynucleotide encoding a native g35030 protein, or alternatively a mutant g35030 protein.

In a second preferred embodiment, these transgenic animals may express a desired polypeptide of interest under the control of regulatory polynucleotides which lead to good  
20 yields in the synthesis of this protein of interest, and optionally a tissue specific expression of this protein of interest.

The design of the transgenic animals of the invention may be made according to the conventional techniques well known from the one skilled in the art. For more details regarding the production of transgenic animals, and specifically transgenic mice, it may be referred to US  
25 Patents Nos 4,873,191, issued Oct. 10, 1989; 5,464,764 issued Nov 7, 1995; and 5,789,215, issued Aug 4, 1998.

Transgenic animals of the present invention are produced by the application of procedures which result in an animal with a genome that has incorporated exogenous genetic material. The procedure involves obtaining the genetic material, or a portion thereof, which  
30 encodes either a g35030 polynucleotide or antisense polynucleotide such as described in the present specification.

A recombinant polynucleotide of the invention is inserted into an embryonic or ES stem cell line. The insertion is preferably made using electroporation, such as described by Thomas et al.(1987). The cells subjected to electroporation are screened (e.g. by selection via selectable  
35 markers, by PCR or by Southern blot analysis) to find positive cells which have integrated the

exogenous recombinant polynucleotide into their genome, preferably via an homologous recombination event. An illustrative positive-negative selection procedure that may be used according to the invention is described by Mansour et al.(1988).

5 Then, the positive cells are isolated, cloned and injected into 3.5 days old blastocysts from mice, such as described by Bradley (1987). The blastocysts are then inserted into a female host animal and allowed to grow to term.

Alternatively, the positive ES cells are brought into contact with embryos at the 2.5 days old 8-16 cell stage (morulae) such as described by Wood et al.(1993) or by Nagy et al.(1993), the ES cells being internalized to colonize extensively the blastocyst including the  
10 cells which will give rise to the germ line.

The offspring of the female host are tested to determine which animals are transgenic e.g. include the inserted exogenous DNA sequence and which are wild-type.

Thus, the present invention also concerns a transgenic animal containing a nucleic acid, a recombinant expression vector or a recombinant host cell according to the invention.

#### 15 **Recombinant Cell Lines Derived From The Transgenic Animals Of The Invention.**

A further object of the invention comprises recombinant host cells obtained from a transgenic animal described herein. In one embodiment the invention encompasses cells derived from non-human host mammals and animals comprising a recombinant vector of the invention or a gene comprising a g35030 nucleic acid sequence disrupted by homologous  
20 recombination with a knock out vector.

Recombinant cell lines may be established *in vitro* from cells obtained from any tissue of a transgenic animal according to the invention, for example by transfection of primary cell cultures with vectors expressing *onc*-genes such as SV40 large T antigen, as described by Chou (1989) and Shay et al.(1991).

25

#### **Assays For Identification Of Compounds For Treatment Of Schizophrenia And Bipolar Disorder**

The present invention provides assays which may be used to test compounds for their ability to treat CNS disorders, and in particular, to ameliorate symptoms of a CNS disorder  
30 mediated by g35030. In preferred embodiments, compounds tested for their ability to ameliorate symptoms of schizophrenia or bipolar disorder mediated by g35030. Compounds may also be tested for their ability to treat related disorders, including among others psychotic disorders, mood disorders, autism, substance dependence and alcoholism, mental retardation, and other psychiatric diseases including cognitive, anxiety, eating, impulse-control, and personality  
35 disorders, as defined with the Diagnosis and Statistical Manual of Mental Disorders fourth edition

(DSM-IV) classification.

The present invention also provides cell and animal, including primate and mouse, models of schizophrenia, bipolar disorder and related disorders.

5 In one aspect, provided are non-cell based, cell based and animal based assays for the identification of such compounds that affect g35030 activity. g35030 activity may be affected by any mechanism; in certain embodiments, g35030 activity is affected by modulating g35030 gene expression or the activity of the g35030 gene product.

10 The present methods allow the identification of compounds that affect g35030 activity directly or indirectly. Thus, the non-cell based, cell based and animal assays of the present invention may also be used to identify compounds that act on an element of a g35030 pathway other than g35030 itself. These compounds can then be used as a therapeutic treatment to modulate g35030 and other gene products involved in schizophrenia, bipolar disorder and related disorders.

*Cell and non-cell based assays*

15 In one aspect, cell based assays using recombinant or non-recombinant cells may be used to identify compounds which modulate g35030 activity.

20 In one aspect, a cell based assay of the invention encompasses a method for identifying a test compound for the treatment of schizophrenia or bipolar disorder comprising (a) exposing a cell to a test compound at a concentration and time sufficient to ameliorate an endpoint related to schizophrenia or bipolar disorder, and (b) determining the level of g35030 activity in a cell. g35030 activity can be measured, for example, by assaying a cell for mRNA transcript level, g35030 peptide expression, localization or protein activity. Preferably the test compound is a compound capable of or suspected to be capable of ameliorating a symptom of schizophrenia, bipolar disorder or a related disorder. Test compounds capable of modulating g35030 activity  
25 may be selected for use in developing medicaments. Such cell based assays are further described herein in the section titled "Method For Screening Ligands That Modulate The Expression Of The g35030 Gene."

30 In another aspect, a cell based assay of the invention encompasses a method for identifying a compound for the treatment of schizophrenia or bipolar disorder comprising (a) exposing a cell to a level of g35030 activity sufficient to cause a schizophrenia-related or bipolar disorder-related endpoint, and (b) exposing said cell to a test compound. A test compound can then be selected according to its ability to ameliorate said schizophrenia-related or bipolar disorder-related endpoints. g35030 activity may be provided by any suitable method, including but not limited to providing a vector containing a g35030 nucleotide sequence,  
35 treating said cell with a compound capable of increasing g35030 expression, and treating said

cell with a g35030 peptide. Preferably said cell is treated with a g35030 peptide comprising a contiguous span of at least 4 amino acids of SEQ ID Nos. 18 to 23. Preferably the test compound is a compound capable of or suspected to be capable of ameliorating a symptom of schizophrenia, bipolar disorder or a related disorder; alternatively, the test compound is suspected of exacerbating an endpoint schizophrenia, bipolar disorder or a related disorder. A test compound capable of ameliorating any detectable symptom or endpoint of a schizophrenia, bipolar disorder or a related disorder may be selected for use in developing medicaments.

In another embodiment, the invention provides cell and non-cell based assays to g35030 to determine whether g35030 peptides bind to the cell surface, and to identify compounds for the treatment of schizophrenia, bipolar disorder and related disorders that interact with a g35030 receptor. In one such embodiment, a g35030 polynucleotide, or fragments thereof, is cloned into expression vectors such as those described herein. The proteins are purified by size, charge, immunochromatography or other techniques familiar to those skilled in the art. Following purification, the proteins are labeled using techniques known to those skilled in the art. The labeled proteins are incubated with cells or cell lines derived from a variety of organs or tissues to allow the proteins to bind to any receptor present on the cell surface. Following the incubation, the cells are washed to remove non-specifically bound protein. The labeled proteins are detected by autoradiography. Alternatively, unlabeled proteins may be incubated with the cells and detected with antibodies having a detectable label, such as a fluorescent molecule, attached thereto. Specificity of cell surface binding may be analyzed by conducting a competition analysis in which various amounts of unlabeled protein are incubated along with the labeled protein. The amount of labeled protein bound to the cell surface decreases as the amount of competitive unlabeled protein increases. As a control, various amounts of an unlabeled protein unrelated to the labeled protein is included in some binding reactions. The amount of labeled protein bound to the cell surface does not decrease in binding reactions containing increasing amounts of unrelated unlabeled protein, indicating that the protein encoded by the nucleic acid binds specifically to the cell surface.

In another embodiment, the present invention comprises non-cell based binding assays, wherein a g35030 polypeptide is prepared and purified as in cell based binding assays described above. Following purification, the proteins are labeled and incubated with a cell membrane extract or isolate derived from any desired cells from any organs, tissue or combination of organs or tissues of interest to allow the g35030 polypeptide to bind to any receptor present on a membrane. Following the incubation, the membranes are washed to remove non-specifically bound protein. The labeled proteins may be detected by autoradiography. Specificity of membrane binding of g35030 may be analyzed by conducting a competition analysis in which

various amounts of a test compound are incubated along with the labeled protein. Any desired test compound, including test polypeptides, can be incubated with the cells. The test compounds may be detected with antibodies having a detectable label, such as a fluorescent molecule, attached thereto. The amount of labeled g35030 polypeptide bound to the cell surface decreases as the amount of competitive test compound increases. As a control, various amounts of an unlabeled protein or a compound unrelated to the test compound is included in some binding reactions. Test compounds capable of reducing the amount of g35030 bound to cell membranes may be selected as a candidate therapeutic compound.

Said cell based assays may comprise cells of any suitable origin; particularly preferred cells are human cells, primate cells, non-human primate cells and mouse cells. If non-human primate cells are used, the g35030 may comprise a nucleotide sequence or be encoded by a nucleotide sequence according to the primate nucleic acid sequences of SEQ ID No. 79 to 132, or a sequence complementary thereto or a fragment thereof.

*Animal model based assay*

Non-human animal based assays may also be used to identify compounds which modulate g35030 activity. The invention encompasses animal models and animal based assays, including non-transgenic or transgenic animals, including animals containing a human or altered form of the g35030 gene.

Thus, the present invention comprises treating an animal affected by schizophrenia or bipolar disorder or symptoms thereof with a test compound capable of directly or indirectly modulating g35030 activity.

In one aspect, an animal based assay of the invention encompasses a method for identifying a test compound for the treatment of schizophrenia or bipolar disorder comprising (a) exposing an animal to a test compound at a concentration and time sufficient to ameliorate an endpoint related to schizophrenia or bipolar disorder, and (b) determining the level of g35030 activity at a site in said animal. g35030 activity can be measured in any suitable cell, tissue or site. Preferably the test compound is a compound capable of or suspected to be capable of ameliorating a symptom of schizophrenia, bipolar disorder or a related disorder. Optionally said test compound is capable or suspected to be capable of modulating g35030 activity. Test compounds capable of modulating g35030 activity may be selected for use in developing medicaments.

In another aspect, an animal based assay of the invention encompasses a method for identifying a compound for the treatment of schizophrenia or bipolar disorder comprising (a) exposing an animal to a level of g35030 activity sufficient to cause a schizophrenia-related or bipolar disorder-related symptom or endpoint, and (b) exposing said animal to a test compound.



A test compound can then be selected according to its ability to ameliorate said schizophrenia-related or bipolar disorder-related endpoints. g35030 activity may be provided by any suitable method, including but not limited to providing a vector containing a g35030 nucleotide sequence, treating said animal with a compound capable of increasing g35030 expression and  
5 treating said cell with a g35030 peptide. Preferably, said animal is treated with a g35030 peptide comprising a contiguous span of at least 4 amino acids of SEQ ID Nos. 18 to 23. Preferably the test compound is a compound capable of or suspected to be capable of ameliorating a symptom of schizophrenia, bipolar disorder or a related disorder; alternatively, the test compound is suspected of exacerbating a symptom of schizophrenia, bipolar disorder or  
10 a related disorder. A test compound capable of ameliorating any detectable symptom or endpoint of a schizophrenia, bipolar disorder or a related disorder may be selected for use in developing medicaments.

Any suitable animal may be used. Preferably, said animal is a primate, a non-human primate, a mammal, or a mouse.

15 In one embodiment, a mouse is treated with a g35030 peptide, exposed to a test compound, and symptoms indicative of schizophrenia, bipolar disorder or a related disorder are assessed by observing stereotypy. In other embodiments, said symptoms are assessed by performing at least one test from the group consisting of home cage observation, neurological evaluation, stress-induced hypothermia, forced swim, PTZ seizure, locomotor activity, tail  
20 suspension, elevated plus maze, novel object recognition, prepulse inhibition, thermal pain, Y-maze, and metabolic chamber tests (Psychoscreen™ tests available from Psychogenics Inc.). Other tests are known in Crawley et al, Horm. Behav. 31(3):197-211 (1997); Crawley, Brain Res 835(1):18-26 (1999) for example.

Any suitable test compound may be used with the screening methods of the invention.  
25 Examples of compounds that may be screened by the methods of the present invention include small organic or inorganic molecules, nucleic acids, including polynucleotides from random and directed polynucleotide libraries, peptides, including peptides derived from random and directed peptide libraries, soluble peptides, fusion peptides, and phosphopeptides, antibodies including polyclonal, monoclonal, chimeric, humanized, and anti-idiotypic antibodies, and single chain  
30 antibodies, FAb, F(ab')<sub>2</sub> and FAb expression library fragments, and epitope-binding fragments thereof. In certain aspects, a compound capable of ameliorating or exacerbating a symptom or endpoint of schizophrenia, bipolar disorder or a related disorder may include, by way of example, antipsychotic drugs in general, neuroleptics, atypical neuroleptics, antidepressants, anti-anxiety drugs, noradrenergic agonists and antagonists, dopaminergic agonists and  
35 antagonists, serotonin reuptake inhibitors, benzodiazepines.

### Methods for screening substances interacting with a g35030 polypeptides

For the purpose of the present invention, a ligand means a molecule, such as a protein, a peptide, an antibody or any synthetic chemical compound capable of binding to the g35030 protein or one of its fragments or variants or to modulate the expression of the polynucleotide coding for the g35030 or a fragment or variant thereof.

In the ligand screening method according to the present invention, a biological sample or a defined molecule to be tested as a putative ligand of the g35030 protein is brought into contact with the corresponding purified g35030 protein, for example the corresponding purified recombinant g35030 protein produced by a recombinant cell host as described hereinbefore, in order to form a complex between this protein and the putative ligand molecule to be tested.

As an illustrative example, to study the interaction of the g35030 protein, or a fragment comprising a contiguous span of at least 4 amino acids, preferably at least 6, or preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID Nos 18 to 23, with drugs or small molecules, such as molecules generated through combinatorial chemistry approaches, the microdialysis coupled to HPLC method described by Wang et al. (1997) or the affinity capillary electrophoresis method described by Bush et al. (1997), can be used.

In further methods, peptides, drugs, fatty acids, lipoproteins, or small molecules which interact with the g35030 protein, or a fragment comprising a contiguous span of at least 4 amino acids, preferably at least 6, or preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID Nos 18 to 23, may be identified using assays such as the following. The molecule to be tested for binding is labeled with a detectable label, such as a fluorescent, radioactive, or enzymatic tag and placed in contact with immobilized g35030 protein, or a fragment thereof under conditions which permit specific binding to occur. After removal of non-specifically bound molecules, bound molecules are detected using appropriate means.

Another object of the present invention comprises methods and kits for the screening of candidate substances that interact with a g35030 polypeptide.

The present invention pertains to methods for screening substances of interest that interact with a g35030 protein or one fragment or variant thereof. By their capacity to bind covalently or non-covalently to a g35030 protein or to a fragment or variant thereof, these substances or molecules may be advantageously used both *in vitro* and *in vivo*.

*In vitro*, said interacting molecules may be used as detection means in order to identify the presence of a g35030 protein in a sample, preferably a biological sample.

A method for the screening of a candidate substance comprises the following steps :

a) providing a polypeptide comprising, consisting essentially of, or consisting of a g35030 protein or a fragment comprising a contiguous span of at least 4 amino acids, preferably at least 6 amino acids, more preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID Nos. 18 to 23;

- 5           b) obtaining a candidate substance;
- c) bringing into contact said polypeptide with said candidate substance; and
- d) detecting the complexes formed between said polypeptide and said candidate substance.

10           The invention further concerns a kit for the screening of a candidate substance interacting with the g35030 polypeptide, wherein said kit comprises:

- a) a g35030 protein having an amino acid sequence selected from the group consisting of the amino acid sequences of SEQ ID Nos. 18 to 23 or a peptide fragment comprising a contiguous span of at least 4 amino acids, preferably at least 6 amino acids, more preferably at least 8 to 10 amino acids, and more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID Nos. 18 to 23; and
- 15           b) optionally means useful to detect the complex formed between the g35030 protein or a peptide fragment or a variant thereof and the candidate substance.

          In a preferred embodiment of the kit described above, the detection means comprise monoclonal or polyclonal antibodies directed against the g35030 protein or a peptide fragment or a variant thereof.

20

          Various candidate substances or molecules can be assayed for interaction with a g35030 polypeptide. These substances or molecules include, without being limited to, natural or synthetic organic compounds or molecules of biological origin such as polypeptides. When the candidate substance or molecule comprise a polypeptide, this polypeptide may be the resulting expression product of a phage clone belonging to a phage-based random peptide library, or alternatively the polypeptide may be the resulting expression product of a cDNA library cloned in a vector suitable for performing a two-hybrid screening assay.

25

          The invention also pertains to kits useful for performing the hereinbefore described screening method. Preferably, such kits comprise a g35030 polypeptide or a fragment or a variant thereof, and optionally means useful to detect the complex formed between the g35030 polypeptide or its fragment or variant and the candidate substance. In a preferred embodiment the detection means comprise monoclonal or polyclonal antibodies directed against the corresponding g35030 polypeptide or a fragment or a variant thereof.

30

#### A. Candidate ligands obtained from random peptide libraries

In a particular embodiment of the screening method, the putative ligand is the expression product of a DNA insert contained in a phage vector (Parmley and Smith, 1988). Specifically, random peptide phages libraries are used. The random DNA inserts encode for peptides of 8 to 20 amino acids in length (Oldenburg K.R. et al., 1992; Valadon P., et al., 1996; Lucas A.H., 1994; Westerink M.A.J., 1995; Felici F. et al., 1991). According to this particular embodiment, the recombinant phages expressing a protein that binds to the immobilized g35030 protein is retained and the complex formed between the g35030 protein and the recombinant phage may be subsequently immunoprecipitated by a polyclonal or a monoclonal antibody directed against the g35030 protein.

Once the ligand library in recombinant phages has been constructed, the phage population is brought into contact with the immobilized g35030 protein. Then the preparation of complexes is washed in order to remove the non-specifically bound recombinant phages. The phages that bind specifically to the g35030 protein are then eluted by a buffer (acid pH) or immunoprecipitated by the monoclonal antibody produced by the hybridoma anti- g35030, and this phage population is subsequently amplified by an over-infection of bacteria (for example E. coli). The selection step may be repeated several times, preferably 2-4 times, in order to select the more specific recombinant phage clones. The last step comprises characterizing the peptide produced by the selected recombinant phage clones either by expression in infected bacteria and isolation, expressing the phage insert in another host-vector system, or sequencing the insert contained in the selected recombinant phages.

#### B. Candidate ligands obtained by competition experiments.

Alternatively, peptides, drugs or small molecules which bind to the g35030 protein, or a fragment comprising a contiguous span of at least 4 amino acids, preferably at least 6 amino acids, more preferably at least 8 to 10 amino acids, and more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID Nos. 18 to 23, may be identified in competition experiments. In such assays, the g35030 protein, or a fragment thereof, is immobilized to a surface, such as a plastic plate. Increasing amounts of the peptides, drugs or small molecules are placed in contact with the immobilized g35030 protein, or a fragment thereof, in the presence of a detectable labeled known g35030 protein ligand. For example, the g35030 ligand may be detectably labeled with a fluorescent, radioactive, or enzymatic tag. The ability of the test molecule to bind the g35030 protein, or a fragment thereof, is determined by measuring the amount of detectably labeled known ligand bound in the presence of the test molecule. A decrease in the amount of known ligand bound to the g35030 protein, or a fragment thereof,

when the test molecule is present indicated that the test molecule is able to bind to the g35030 protein, or a fragment thereof.

### C. Candidate ligands obtained by affinity chromatography.

Proteins or other molecules interacting with the g35030 protein, or a fragment  
5 comprising a contiguous span of at 4 amino acids, preferably at least 6 amino acids, more preferably at least 8 to 10 amino acids, and more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID Nos 18 to 23, can also be found using affinity columns which contain the g35030 protein, or a fragment thereof. The g35030 protein, or a fragment thereof, may be attached to the column using conventional techniques including chemical coupling to a  
10 suitable column matrix such as agarose, Affi Gel®, or other matrices familiar to those of skill in art. In some embodiments of this method, the affinity column contains chimeric proteins in which the g35030 protein, or a fragment thereof, is fused to glutathion S transferase (GST). A mixture of cellular proteins or pool of expressed proteins as described above is applied to the affinity column. Proteins or other molecules interacting with the g35030 protein, or a fragment  
15 thereof, attached to the column can then be isolated and analyzed on 2-D electrophoresis gel as described in Ramunsen et al. (1997). Alternatively, the proteins retained on the affinity column can be purified by electrophoresis based methods and sequenced. The same method can be used to isolate antibodies, to screen phage display products, or to screen phage display human antibodies.

### 20 D. Candidate ligands obtained by optical biosensor methods

Proteins interacting with the g35030 protein, or a fragment comprising a contiguous span of at least 4 amino acids, preferably at least 6 amino acids, more preferably at least 8 to 10 amino acids, and more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ  
25 ID Nos. 18 to 23, can also be screened by using an Optical Biosensor as described in Edwards and Leatherbarrow (1997) and also in Szabo et al. (1995). This technique permits the detection of interactions between molecules in real time, without the need of labeled molecules. This technique is based on the surface plasmon resonance (SPR) phenomenon. Briefly, the candidate ligand molecule to be tested is attached to a surface (such as a carboxymethyl dextran matrix). A light beam is directed towards the side of the surface that does not contain the sample to be  
30 tested and is reflected by said surface. The SPR phenomenon causes a decrease in the intensity of the reflected light with a specific association of angle and wavelength. The binding of candidate ligand molecules cause a change in the refraction index on the surface, which change is detected as a change in the SPR signal. For screening of candidate ligand molecules or substances that are able to interact with the g35030 protein, or a fragment thereof, the g35030

protein, or a fragment thereof, is immobilized onto a surface. This surface comprises one side of a cell through which flows the candidate molecule to be assayed. The binding of the candidate molecule on the g35030 protein, or a fragment thereof, is detected as a change of the SPR signal. The candidate molecules tested may be proteins, peptides, carbohydrates, lipids, or small molecules generated by combinatorial chemistry. This technique may also be performed by immobilizing eukaryotic or prokaryotic cells or lipid vesicles exhibiting an endogenous or a recombinantly expressed g35030 protein at their surface.

The main advantage of the method is that it allows the determination of the association rate between the g35030 protein and molecules interacting with the g35030 protein. It is thus possible to select specifically ligand molecules interacting with the g35030 protein, or a fragment thereof, through strong or conversely weak association constants.

#### **E. Candidate ligands obtained through a two-hybrid screening assay.**

The yeast two-hybrid system is designed to study protein-protein interactions *in vivo* (Fields and Song, 1989), and relies upon the fusion of a bait protein to the DNA binding domain of the yeast Gal4 protein. This technique is also described in the US Patent N° US 5,667,973 and the US Patent N° 5,283,173 (Fields et al.).

The general procedure of library screening by the two-hybrid assay may be performed as described by Harper et al. (1993) or as described by Cho et al. (1998) or also Fromont-Racine et al. (1997).

The bait protein or polypeptide comprises, consists essentially of, or consists of a g35030 polypeptide or a fragment comprising a contiguous span of at least 4 amino acids, preferably at least 6 amino acids, more preferably at least 8 to 10 amino acids, and more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID Nos. 18 to 23.

More precisely, the nucleotide sequence encoding the g35030 polypeptide or a fragment or variant thereof is fused to a polynucleotide encoding the DNA binding domain of the GAL4 protein, the fused nucleotide sequence being inserted in a suitable expression vector, for example pAS2 or pM3.

Then, a human cDNA library is constructed in a specially designed vector, such that the human cDNA insert is fused to a nucleotide sequence in the vector that encodes the transcriptional domain of the GAL4 protein. Preferably, the vector used is the pACT vector. The polypeptides encoded by the nucleotide inserts of the human cDNA library are termed "prey" polypeptides.

A third vector contains a detectable marker gene, such as beta galactosidase gene or CAT gene that is placed under the control of a regulation sequence that is responsive to the

binding of a complete Gal4 protein containing both the transcriptional activation domain and the DNA binding domain. For example, the vector pG5EC may be used.

Two different yeast strains are also used. As an illustrative but non limiting example the two different yeast strains may be the followings :

- 5       - Y190, the phenotype of which is (*MATa, Leu2-3, 112 ura3-12, trp1-901, his3-D200, ade2-101, gal4Dgal180D URA3 GAL-LacZ, LYS GAL-HIS3, cyh*);
- Y187, the phenotype of which is (*MATa gal4 gal80 his3 trp1-901 ade2-101 ura3-52 leu2-3, -112 URA3 GAL-lacZmer*), which is the opposite mating type of Y190.

10       Briefly, 20 µg of pAS2/ g35030 and 20 µg of pACT-cDNA library are co-transformed into yeast strain Y190. The transformants are selected for growth on minimal media lacking histidine, leucine and tryptophan, but containing the histidine synthesis inhibitor 3-AT (50 mM). Positive colonies are screened for beta galactosidase by filter lift assay. The double positive colonies (*His<sup>+</sup>, beta-gal<sup>+</sup>*) are then grown on plates lacking histidine, leucine, but containing tryptophan and cycloheximide (10 mg/ml) to select for loss of pAS2/ g35030

15       plasmids by retention of pACT-cDNA library plasmids. The resulting Y190 strains are mated with Y187 strains expressing g35030 or non-related control proteins; such as cyclophilin B, lamin, or SNF1, as *Gal4* fusions as described by Harper et al. (1993) and by Bram et al. (Bram RJ et al., 1993), and screened for beta galactosidase by filter lift assay. Yeast clones that are *beta gal-* after mating with the control *Gal4* fusions are considered false positives.

20       In another embodiment of the two-hybrid method according to the invention, interaction between the g35030 or a fragment or variant thereof with cellular proteins may be assessed using the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech). As described in the manual accompanying the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech), nucleic acids encoding the g35030 protein or a portion thereof, are inserted into an

25       expression vector such that they are in frame with DNA encoding the DNA binding domain of the yeast transcriptional activator GAL4. A desired cDNA, preferably human cDNA, is inserted into a second expression vector such that they are in frame with DNA encoding the activation domain of GAL4. The two expression plasmids are transformed into yeast and the yeast are plated on selection medium which selects for expression of selectable markers on each of the expression

30       vectors as well as GAL4 dependent expression of the HIS3 gene. Transformants capable of growing on medium lacking histidine are screened for GAL4 dependent lacZ expression. Those cells which are positive in both the histidine selection and the lacZ assay contain interaction between g35030 and the protein or peptide encoded by the initially selected cDNA insert.

**Method For Screening Substances Interacting With The Regulatory Sequences Of A g35030 Gene.**

The present invention also concerns a method for screening substances or molecules that are able to interact with the regulatory sequences of the g35030 gene, such as for example promoter or enhancer sequences.

Nucleic acids encoding proteins which are able to interact with the regulatory sequences of the g35030 gene, more particularly a nucleotide sequence selected from the group consisting of the polynucleotides of the 5' and 3' regulatory region or a fragment or variant thereof, and preferably a variant comprising one of the biallelic markers of the invention, may be identified by using a one-hybrid system, such as that described in the booklet enclosed in the Matchmaker One-Hybrid System kit from Clontech (Catalog Ref. n° K1603-1). Briefly, the target nucleotide sequence is cloned upstream of a selectable reporter sequence and the resulting DNA construct is integrated in the yeast genome (*Saccharomyces cerevisiae*). The yeast cells containing the reporter sequence in their genome are then transformed with a library comprising fusion molecules between cDNAs encoding candidate proteins for binding onto the regulatory sequences of the g35030 gene and sequences encoding the activator domain of a yeast transcription factor such as GAL4. The recombinant yeast cells are plated in a culture broth for selecting cells expressing the reporter sequence. The recombinant yeast cells thus selected contain a fusion protein that is able to bind onto the target regulatory sequence of the g35030 gene. Then, the cDNAs encoding the fusion proteins are sequenced and may be cloned into expression or transcription vectors *in vitro*. The binding of the encoded polypeptides to the target regulatory sequences of the g35030 gene may be confirmed by techniques familiar to the one skilled in the art, such as gel retardation assays or DNase protection assays.

Gel retardation assays may also be performed independently in order to screen candidate molecules that are able to interact with the regulatory sequences of the g35030 gene, such as described by Fried and Crothers (1981), Garner and Revzin (1981) and Dent and Latchman (1993). These techniques are based on the principle according to which a DNA fragment which is bound to a protein migrates slower than the same unbound DNA fragment. Briefly, the target nucleotide sequence is labeled. Then the labeled target nucleotide sequence is brought into contact with either a total nuclear extract from cells containing transcription factors, or with different candidate molecules to be tested. The interaction between the target regulatory sequence of the g35030 gene and the candidate molecule or the transcription factor is detected after gel or capillary electrophoresis through a retardation in the migration.

**Method For Screening Ligands That Modulate The Expression Of The g35030**



### Gene

Another subject of the present invention is a method for screening molecules that modulate the expression of the g35030 protein. Such a screening method comprises the steps of:

- 5           a) cultivating a prokaryotic or an eukaryotic cell that has been transfected with a nucleotide sequence encoding the g35030 protein or a variant or a fragment thereof, placed under the control of its own promoter;
- b) bringing into contact the cultivated cell with a molecule to be tested;
- c) quantifying the expression of the g35030 protein or a variant or a fragment thereof.

10           In an embodiment, the nucleotide sequence encoding the g35030 protein or a variant or a fragment thereof comprises an allele of at least one g35030 related biallelic marker.

            Using DNA recombination techniques well known by the one skilled in the art, the g35030 protein encoding DNA sequence is inserted into an expression vector, downstream from its promoter sequence. As an illustrative example, the promoter sequence of the g35030 gene is contained in the nucleic acid of the 5' regulatory region.

15           The quantification of the expression of the g35030 protein may be realized either at the mRNA level or at the protein level. In the latter case, polyclonal or monoclonal antibodies may be used to quantify the amounts of the g35030 protein that have been produced, for example in an ELISA or a RIA assay.

20           In a preferred embodiment, the quantification of the g35030 mRNA is realized by a quantitative PCR amplification of the cDNA obtained by a reverse transcription of the total mRNA of the cultivated g35030 -transfected host cell, using a pair of primers specific for g35030.

25           The present invention also concerns a method for screening substances or molecules that are able to increase, or in contrast to decrease, the level of expression of the g35030 gene. Such a method may allow the one skilled in the art to select substances exerting a regulating effect on the expression level of the g35030 gene and which may be useful as active ingredients included in pharmaceutical compositions for treating patients suffering from diseases.

30           Thus, is also part of the present invention a method for screening of a candidate substance or molecule that modulated the expression of the g35030 gene, this method comprises the following steps:

- providing a recombinant cell host containing a nucleic acid, wherein said nucleic acid comprises a nucleotide sequence of the 5' regulatory region or a biologically active fragment or variant thereof located upstream a polynucleotide encoding a detectable protein;
- 35           - obtaining a candidate substance; and

- determining the ability of the candidate substance to modulate the expression levels of the polynucleotide encoding the detectable protein.

In a further embodiment, the nucleic acid comprising the nucleotide sequence of the 5' regulatory region or a biologically active fragment or variant thereof also includes a 5'UTR region of the g35030 cDNA of SEQ ID No 2 to 17, or one of its biologically active fragments or variants thereof.

Among the preferred polynucleotides encoding a detectable protein, there may be cited polynucleotides encoding beta galactosidase, green fluorescent protein (GFP) and chloramphenicol acetyl transferase (CAT).

The invention also pertains to kits useful for performing the herein described screening method. Preferably, such kits comprise a recombinant vector that allows the expression of a nucleotide sequence of the 5' regulatory region or a biologically active fragment or variant thereof located upstream and operably linked to a polynucleotide encoding a detectable protein or the g35030 protein or a fragment or a variant thereof.

In another embodiment of a method for the screening of a candidate substance or molecule that modulates the expression of the g35030 gene, wherein said method comprises the following steps:

a) providing a recombinant host cell containing a nucleic acid, wherein said nucleic acid comprises a 5'UTR sequence of a g35030 cDNA, preferably of a g35030, or one of its biologically active fragments or variants, the 5'UTR sequence or its biologically active fragment or variant being operably linked to a polynucleotide encoding a detectable protein;

b) obtaining a candidate substance; and

c) determining the ability of the candidate substance to modulate the expression levels of the polynucleotide encoding the detectable protein.

In a specific embodiment of the above screening method, the nucleic acid that comprises a nucleotide sequence selected from the group consisting of the 5'UTR sequence of a g35030 cDNA, preferably of a g35030 or one of its biologically active fragments or variants, includes a promoter sequence which is endogenous with respect to the g35030 5'UTR sequence.

In another specific embodiment of the above screening method, the nucleic acid that comprises a nucleotide sequence selected from the group consisting of the 5'UTR sequence of a g35030 cDNA or one of its biologically active fragments or variants, includes a promoter sequence which is exogenous with respect to the g35030 5'UTR sequence defined therein.

In a further preferred embodiment, the nucleic acid comprising the 5'-UTR sequence of a g35030 cDNA or the biologically active fragments thereof includes a g35030-related biallelic marker.

The invention further comprises a kit for the screening of a candidate substance modulating the expression of the g35030 gene, wherein said kit comprises a recombinant vector that comprises a nucleic acid including a 5'UTR sequence of the g35030 cDNA of SEQ ID Nos 2 to 17, or one of their biologically active fragments or variants, the 5'UTR sequence or its  
5 biologically active fragment or variant being operably linked to a polynucleotide encoding a detectable protein.

For the design of suitable recombinant vectors useful for performing the screening methods described above, it will be referred to the section of the present specification wherein the preferred recombinant vectors of the invention are detailed.

10 Expression levels and patterns of g35030 may be analyzed by solution hybridization with long probes as described in International Patent Application No. WO 97/05277. Briefly, the g35030 cDNA or the g35030 genomic DNA described above, or fragments thereof, is inserted at a cloning site immediately downstream of a bacteriophage (T3, T7 or SP6) RNA polymerase promoter to produce antisense RNA. Preferably, the g35030 insert comprises at  
15 least 100 or more consecutive nucleotides of the genomic DNA sequence or the cDNA sequences. The plasmid is linearized and transcribed in the presence of ribonucleotides comprising modified ribonucleotides (i.e. biotin-UTP and DIG-UTP). An excess of this doubly labeled RNA is hybridized in solution with mRNA isolated from cells or tissues of interest. The hybridization is performed under standard stringent conditions (40-50°C for 16 hours in an 80%  
20 formamide, 0.4 M NaCl buffer, pH 7-8). The unhybridized probe is removed by digestion with ribonucleases specific for single-stranded RNA (i.e. RNases CL3, T1, Phy M, U2 or A). The presence of the biotin-UTP modification enables capture of the hybrid on a microtitration plate coated with streptavidin. The presence of the DIG modification enables the hybrid to be detected and quantified by ELISA using an anti-DIG antibody coupled to alkaline phosphatase.

25 Quantitative analysis of g35030 gene expression may also be performed using arrays. As used herein, the term array means a one dimensional, two dimensional, or multidimensional arrangement of a plurality of nucleic acids of sufficient length to permit specific detection of expression of mRNAs capable of hybridizing thereto. For example, the arrays may contain a plurality of nucleic acids derived from genes whose expression levels are to be assessed. The  
30 arrays may include the g35030 genomic DNA, the g35030 cDNA sequences or the sequences complementary thereto or fragments thereof, particularly those comprising at least one of the biallelic markers according to the present invention. Preferably, the fragments are at least 15 nucleotides in length. In other embodiments, the fragments are at least 25 nucleotides in length. In some embodiments, the fragments are at least 50 nucleotides in length. More preferably, the  
35 fragments are at least 100 nucleotides in length. In another preferred embodiment, the

fragments are more than 100 nucleotides in length. In some embodiments the fragments may be more than 500 nucleotides in length.

For example, quantitative analysis of g35030 gene expression may be performed with a complementary DNA microarray as described by Schena et al.(1995 and 1996). Full length  
5 g35030 cDNAs or fragments thereof are amplified by PCR and arrayed from a 96-well microtiter plate onto silylated microscope slides using high-speed robotics. Printed arrays are incubated in a humid chamber to allow rehydration of the array elements and rinsed, once in 0. 2% SDS for 1 min, twice in water for 1 min and once for 5 min in sodium borohydride solution. The arrays are submerged in water for 2 min at 95°C, transferred into 0. 2% SDS for 1 min,  
10 rinsed twice with water, air dried and stored in the dark at 25°C.

Cell or tissue mRNA is isolated or commercially obtained and probes are prepared by a single round of reverse transcription. Probes are hybridized to 1 cm<sup>2</sup> microarrays under a 14 x 14 mm glass coverslip for 6-12 hours at 60°C. Arrays are washed for 5 min at 25°C in low stringency wash buffer (1 x SSC/0. 2% SDS), then for 10 min at room temperature in high  
15 stringency wash buffer (0. 1 x SSC/0. 2% SDS). Arrays are scanned in 0. 1 x SSC using a fluorescence laser scanning device fitted with a custom filter set. Accurate differential expression measurements are obtained by taking the average of the ratios of two independent hybridizations.

Quantitative analysis of g35030 gene expression may also be performed with full length  
20 g35030 cDNAs or fragments thereof in complementary DNA arrays as described by Pietu et al.(1996). The full length g35030 cDNA or fragments thereof is PCR amplified and spotted on membranes. Then, mRNAs originating from various tissues or cells are labeled with radioactive nucleotides. After hybridization and washing in controlled conditions, the hybridized mRNAs are detected by phospho-imaging or autoradiography. Duplicate  
25 experiments are performed and a quantitative analysis of differentially expressed mRNAs is then performed.

Alternatively, expression analysis using the g35030 genomic DNA, the g35030 cDNA, or fragments thereof can be done through high density nucleotide arrays as described by  
30 Lockhart et al.(1996) and Sosnowsky et al.(1997). Oligonucleotides of 15-50 nucleotides from the sequences of the g35030 genomic DNA, the g35030 cDNA sequences particularly those comprising at least one of biallelic markers according the present invention, or the sequences complementary thereto, are synthesized directly on the chip (Lockhart et al., supra) or synthesized and then addressed to the chip (Sosnowski et al., supra). Preferably, the oligonucleotides are about 20 nucleotides in length.

35 g35030 cDNA probes labeled with an appropriate compound, such as biotin,

digoxigenin or fluorescent dye, are synthesized from the appropriate mRNA population and then randomly fragmented to an average size of 50 to 100 nucleotides. The said probes are then hybridized to the chip. After washing as described in Lockhart et al., supra and application of different electric fields (Sosnowsky et al., 1997), the dyes or labeling compounds are detected and quantified. Duplicate hybridizations are performed. Comparative analysis of the intensity of the signal originating from cDNA probes on the same target oligonucleotide in different cDNA samples indicates a differential expression of g35030 mRNA.

#### Methods For Inhibiting The Expression Of A g35030 Gene

Other therapeutic compositions according to the present invention comprise advantageously an oligonucleotide fragment of the nucleic sequence of g35030 as an antisense tool or a triple helix tool that inhibits the expression of the corresponding g35030 gene. A preferred fragment of the nucleic sequence of g35030 comprises an allele of at least one of the biallelic markers of the invention.

#### Antisense Approach

Preferred methods using antisense polynucleotide according to the present invention are the procedures described by Sczakiel et al.(1995).

Preferably, the antisense tools are chosen among the polynucleotides (15-200 bp long) that are complementary to the 5' end of the g35030 mRNA. In another embodiment, a combination of different antisense polynucleotides complementary to different parts of the desired targeted gene are used.

Preferred antisense polynucleotides according to the present invention are complementary to a sequence of the mRNAs of g35030 that contains either the translation initiation codon ATG or a splicing donor or acceptor site.

The antisense nucleic acids should have a length and melting temperature sufficient to permit formation of an intracellular duplex having sufficient stability to inhibit the expression of the g35030 mRNA in the duplex. Strategies for designing antisense nucleic acids suitable for use in gene therapy are disclosed in Green et al., (1986) and Izant and Weintraub, (1984).

In some strategies, antisense molecules are obtained by reversing the orientation of the g35030 coding region with respect to a promoter so as to transcribe the opposite strand from that which is normally transcribed in the cell. The antisense molecules may be transcribed using in vitro transcription systems such as those which employ T7 or SP6 polymerase to generate the transcript. Another approach involves transcription of g35030 antisense nucleic acids in vivo by operably linking DNA containing the antisense sequence to a promoter in a suitable expression vector.

Alternatively, suitable antisense strategies are those described by Rossi et al.(1991), in the International Applications Nos. WO 94/23026, WO 95/04141, WO 92/18522 and in the European Patent Application No. EP 0 572 287 A2

5 An alternative to the antisense technology that is used according to the present invention comprises using ribozymes that will bind to a target sequence via their complementary polynucleotide tail and that will cleave the corresponding RNA by hydrolyzing its target site (namely "hammerhead ribozymes"). Briefly, the simplified cycle of a hammerhead ribozyme comprises (1) sequence specific binding to the target RNA via complementary antisense sequences; (2) site-specific hydrolysis of the cleavable motif of the target strand; and (3) release of cleavage products, which gives rise to another catalytic cycle. 10 Indeed, the use of long-chain antisense polynucleotide (at least 30 bases long) or ribozymes with long antisense arms are advantageous. A preferred delivery system for antisense ribozyme is achieved by covalently linking these antisense ribozymes to lipophilic groups or to use liposomes as a convenient vector. Preferred antisense ribozymes according to the present 15 invention are prepared as described by Sczakiel et al.(1995).

#### Triple Helix Approach

The g35030 genomic DNA may also be used to inhibit the expression of the g35030 gene based on intracellular triple helix formation.

20 Triple helix oligonucleotides are used to inhibit transcription from a genome. They are particularly useful for studying alterations in cell activity when it is associated with a particular gene.

Similarly, a portion of the g35030 genomic DNA can be used to study the effect of inhibiting g35030 transcription within a cell. Traditionally, homopurine sequences were considered the most useful for triple helix strategies. However, homopyrimidine sequences can 25 also inhibit gene expression. Such homopyrimidine oligonucleotides bind to the major groove at homopurine:homopyrimidine sequences. Thus, both types of sequences from the g35030 genomic DNA are contemplated within the scope of this invention.

To carry out gene therapy strategies using the triple helix approach, the sequences of the g35030 genomic DNA are first scanned to identify 10-mer to 20-mer homopyrimidine or homopurine stretches which could be used in triple-helix based strategies for inhibiting g35030 30 expression. Following identification of candidate homopyrimidine or homopurine stretches, their efficiency in inhibiting g35030 expression is assessed by introducing varying amounts of oligonucleotides containing the candidate sequences into tissue culture cells which express the g35030 gene.

The oligonucleotides can be introduced into the cells using a variety of methods known to those skilled in the art, including but not limited to calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection or native uptake.

5 Treated cells are monitored for altered cell function or reduced g35030 expression using techniques such as Northern blotting, RNase protection assays, or PCR based strategies to monitor the transcription levels of the g35030 gene in cells which have been treated with the oligonucleotide.

10 The oligonucleotides which are effective in inhibiting gene expression in tissue culture cells may then be introduced in vivo using the techniques described above in the antisense approach at a dosage calculated based on the in vitro results, as described in antisense approach.

In some embodiments, the natural (beta) anomers of the oligonucleotide units can be replaced with alpha anomers to render the oligonucleotide more resistant to nucleases. Further, an intercalating agent such as ethidium bromide, or the like, can be attached to the 3' end of the alpha oligonucleotide to stabilize the triple helix. For information on the generation of  
15 oligonucleotides suitable for triple helix formation see Griffin et al.(1989).

### **Pharmaceutical Compositions And Formulations**

#### **G35030-modulating Compounds**

20 Using the methods disclosed herein, compounds that selectively modulate g35030 activity *in vitro* and *in vivo* may be identified. The compounds identified by the process of the invention include, for example, antibodies having binding specificity for the g35030 peptide. It is also expected that homologues of g35030 may be useful for modulating g35030-mediated activity and the related physiological condition associated with schizophrenia or bipolar disorder. Generally, it is further expected that assay methods of the present invention based on the role of g35030 in  
25 central nervous system disorder may be used to identify compounds capable of intervening in the assay cascade of the invention.

#### **Indications**

30 While g35030 has demonstrated an association with schizophrenia and bipolar disorder, indications involving g35030 may include various central nervous system disorders. Nervous system disorders are expected to have complex genetic bases and often share certain symptoms. In particular, as described herein, indications may include schizophrenia and other psychotic disorders, mood disorders, autism, substance dependence and alcoholism, mental retardation, and other psychiatric diseases including cognitive, anxiety, eating, impulse-control, and personality disorders, as defined with the Diagnosis and Statistical Manual of Mental Disorders fourth edition  
35 (DSM-IV) classification.

#### Pharmaceutical Formulations and Routes of Administration

The compounds identified using the methods of the present invention can be administered to a mammal, including a human patient, alone or in pharmaceutical compositions where they are mixed with suitable carriers or excipient(s) at therapeutically effective doses to treat or ameliorate schizophrenia or bipolar disorder related disorders. A therapeutically effective dose further refers to that amount of the compound sufficient to result in amelioration of symptoms as determined by the methods described herein. Preferably, a therapeutically effective dosage is suitable for continued periodic use or administration. Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest edition.

#### Routes of Administration

Suitable routes of administration include oral, rectal, transmucosal, or intestinal administration, parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal or intraocular injections. A particularly useful method of administering compounds for treating central nervous system disease involves surgical implantation of a device for delivering the compound over an extended period of time. Sustained release formulations of the invented medicaments particularly are contemplated.

#### Composition/Formulation

Pharmaceutical compositions and medicaments for use in accordance with the present invention may be formulated in a conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries. Proper formulation is dependent upon the route of administration chosen.

For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer such as a phosphate or bicarbonate buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with fillers such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable



for such administration.

For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebulizer, with the use of a suitable gaseous propellant, *e.g.*, carbon dioxide. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.*, gelatin, for use in an inhaler or insufflator, may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampoules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Aqueous suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Alternatively, the active ingredient may be in powder or lyophilized form for constitution with a suitable vehicle, such as sterile pyrogen-free water, before use.

In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable matrices of solid hydrophobic polymers containing the therapeutic agent. Various sustained release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compounds for a few weeks up to over 100 days.

Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols.

5                   Effective Dosage

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve their intended purpose. More specifically, a therapeutically effective amount means an amount effective to prevent development of or to alleviate the existing symptoms of the subject being  
10                   treated. Determination of the effective amounts is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein.

For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays, and a dose can be formulated in animal models. Such information can be used to more accurately determine useful doses in humans.

15                   A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms in a patient. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD50, (the dose lethal to 50% of the test population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and  
20                   therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD50 and ED50. Compounds which exhibit high therapeutic indices are preferred.

The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50, with little or no toxicity. The  
25                   dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. (*See, e.g., Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1).*

30                   Computer-Related Embodiments

As used herein the term "nucleic acid codes of the invention" encompass the nucleotide sequences comprising, consisting essentially of, or consisting of any one of the following:

a) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, 1000 or 2000 nucleotides of SEQ ID No. 1, and the complements thereof,  
35                   wherein said contiguous span comprises at least one of the following nucleotide positions of

SEQ ID No 1: 201123 to 247802 and 199122 to 249803;

b) a contiguous span of at least 8, 12, 15, 18, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100 or 200 nucleotides, to the extent that such a length is consistent with the particular sequence ID, of SEQ ID Nos. 2 to 17 or the complements thereof;

5 c) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, 1000 or 2000 nucleotides of any of SEQ ID Nos. 24, 25, 28 and 74, and the complements thereof, to the extent that such a length is consistent with the particular sequence ID;

10 d) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, 1000 or 2000 nucleotides of any of SEQ ID Nos. 79 to 132, and the complements thereof, to the extent that such a length is consistent with the particular sequence ID;

e) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90 or 100 nucleotides of SEQ ID No. 1 or the complements thereof wherein said contiguous span comprises at least one of the following nucleotide positions of SEQ ID No 1: 201123 to 201234, 201123 to 201560, 214676 to 214793, 215702 to 215746, 216836 to 216994, 216836 to 217077, 217671 to 217764, 227655 to 227736, 238715 to 238919, 240440 to 240673, 240440 to 241153, 241072 to 241291, 244353 to 244561, 246273 to 247802, 199122 to 201122 and 247803 to 249803;

15 f) a contiguous span according to a), b), c), d) or e), wherein said span includes a biallelic marker;

20 g) a contiguous span according to a), b), c), d) or e), wherein said span includes a biallelic marker selected from the group consisting of A13 to A18, A20 to A46, A49 to A52, A55, A57, A59 to A63, A72 to A73, A76 and A123 ; and

25 h) a contiguous span according to f), wherein said span includes a biallelic marker with the alternative allele present at said biallelic marker;

The "nucleic acid codes of the invention" further encompass nucleotide sequences homologous to a contiguous span of at least 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, 1000 or 2000 nucleotides, to the extent that such a length is consistent with the particular sequence of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132, and the complements thereof.

30 The "nucleic acid codes of the invention" also encompass nucleotide sequences homologous to a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90 or 100 nucleotides of SEQ ID No. 1 or the complements thereof, wherein said contiguous span comprises at least one of the following nucleotide positions of SEQ ID No. 1: 201123 to 201234, 201123 to 201560, 214676 to 214793, 215702 to 215746, 216836 to 216994, 216836 to 217077, 217671 to 217764, 227655 to 227736, 238715 to 238919, 240440 to 240673,

35

240440 to 241153, 241072 to 241291, 244353 to 244561, 246273 to 247802, 201123 to 247802, 199122 to 201122, 247803 to 249803 and 199122 to 249803.

Homologous sequences refer to a sequence having at least 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, or 75% homology to these contiguous spans. Homology may be  
5 determined using any method described herein, including BLAST2N with the default parameters or with any modified parameters. Homologous sequences also may include RNA sequences in which uridines replace the thymines in the nucleic acid codes of the invention. It will be appreciated that the nucleic acid codes of the invention can be represented in the traditional single character format (See the inside back cover of Stryer, Lubert. *Biochemistry*,  
10 3<sup>rd</sup> edition. W. H Freeman & Co., New York.) or in any other format or code which records the identity of the nucleotides in a sequence.

As used herein the term "polypeptide codes of SEQ ID Nos. 18 to 23" encompasses the polypeptide sequence of SEQ ID Nos 18 to 23, polypeptide sequences homologous to the polypeptides of SEQ ID Nos. 18 to 23, or fragments of any of the preceding sequences.  
15 Homologous polypeptide sequences refer to a polypeptide sequence having at least 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, 75% homology to one of the polypeptide sequences of SEQ ID Nos. 18 to 23. Homology may be determined using any of the computer programs and parameters described herein, including FASTA with the default parameters or with any modified parameters. The homologous sequences may be obtained using any of the procedures described herein or may  
20 result from the correction of a sequencing error as described above. The polypeptide fragments comprise at least 4, 6, 8, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids of the polypeptides of SEQ ID Nos. 18 to 23. Preferably, the fragments are novel fragments. It will be appreciated that the polypeptide codes of the SEQ ID Nos. 18 to 23 can be represented in the traditional single character format or three letter format (See the inside back cover of Starrier,  
25 Lubert. *Biochemistry*, 3<sup>rd</sup> edition. W. H Freeman & Co., New York.) or in any other format which relates the identity of the polypeptides in a sequence.

It will be appreciated by those skilled in the art that the nucleic acid codes of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 and polypeptide codes of SEQ ID Nos. 18 to 23 can be stored, recorded, and manipulated on any medium which can be read and accessed by a  
30 computer. As used herein, the words "recorded" and "stored" refer to a process for storing information on a computer medium. A skilled artisan can readily adopt any of the presently known methods for recording information on a computer readable medium to generate embodiment comprising one or more of nucleic acid codes of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132, or one or more of the polypeptide codes of SEQ ID Nos. 18 to 23. Another aspect of the  
35 present invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, 20,

25, 30, or 50 nucleic acid codes of SEQ ID Nos 1 to 17, 24, 25, 28, 74 and 79 to 132. Another aspect of the present invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, 20, 25, 30, or 50 polypeptide codes of SEQ ID Nos 18 to 23.

Computer readable media include magnetically readable media, optically readable media, electronically readable media and magnetic/optical media. For example, the computer readable media may be a hard disk, a floppy disk, a magnetic tape, CD-ROM, Digital Versatile Disk (DVD), Random Access Memory (RAM), or Read Only Memory (ROM) as well as other types of other media known to those skilled in the art.

Embodiments of the present invention include systems, particularly computer systems which store and manipulate the sequence information described herein. One example of a computer system 100 is illustrated in block diagram form in Figure 12. As used herein, "a computer system" refers to the hardware components, software components, and data storage components used to analyze the nucleotide sequences of the nucleic acid codes of SEQ ID Nos 1 to 17, 24, 25, 28, 74 and 79 to 132, or the amino acid sequences of the polypeptide codes of SEQ ID Nos. 18 to 23. In one embodiment, the computer system 100 is a Sun Enterprise 1000 server (Sun Microsystems, Palo Alto, CA). The computer system 100 preferably includes a processor for processing, accessing and manipulating the sequence data. The processor 105 can be any well-known type of central processing unit, such as the Pentium III from Intel Corporation, or similar processor from Sun, Motorola, Compaq or International Business Machines.

Preferably, the computer system 100 is a general purpose system that comprises the processor 105 and one or more internal data storage components 110 for storing data, and one or more data retrieving devices for retrieving the data stored on the data storage components. A skilled artisan can readily appreciate that any one of the currently available computer systems are suitable.

In one particular embodiment, the computer system 100 includes a processor 105 connected to a bus which is connected to a main memory 115 (preferably implemented as RAM) and one or more internal data storage devices 110, such as a hard drive and/or other computer readable media having data recorded thereon. In some embodiments, the computer system 100 further includes one or more data retrieving device 118 for reading the data stored on the internal data storage devices 110.

The data retrieving device 118 may represent, for example, a floppy disk drive, a compact disk drive, a magnetic tape drive, etc. In some embodiments, the internal data storage device 110 is a removable computer readable medium such as a floppy disk, a compact disk, a magnetic tape, etc. containing control logic and/or data recorded thereon. The computer system 100 may advantageously include or be programmed by appropriate software for reading the control logic

and/or the data from the data storage component once inserted in the data retrieving device.

The computer system 100 includes a display 120 which is used to display output to a computer user. It should also be noted that the computer system 100 can be linked to other computer systems 125a-c in a network or wide area network to provide centralized access to the computer system 100.

Software for accessing and processing the nucleotide sequences of the nucleic acid codes of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132, or the amino acid sequences of the polypeptide codes of SEQ ID Nos. 18 to 23 (such as search tools, compare tools, and modeling tools etc.) may reside in main memory 115 during execution.

In some embodiments, the computer system 100 may further comprise a sequence comparer for comparing the above-described nucleic acid codes of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 or polypeptide codes of SEQ ID Nos. 18 to 23 stored on a computer readable medium to reference nucleotide or polypeptide sequences stored on a computer readable medium. A "sequence comparer" refers to one or more programs which are implemented on the computer system 100 to compare a nucleotide or polypeptide sequence with other nucleotide or polypeptide sequences and/or compounds including but not limited to peptides, peptidomimetics, and chemicals stored within the data storage means. For example, the sequence comparer may compare the nucleotide sequences of the nucleic acid codes of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132, or the amino acid sequences of the polypeptide codes of SEQ ID Nos. 18 to 23 stored on a computer readable medium to reference sequences stored on a computer readable medium to identify homologies, motifs implicated in biological function, or structural motifs. The various sequence comparer programs identified elsewhere in this patent specification are particularly contemplated for use in this aspect of the invention.

Figure 13 is a flow diagram illustrating one embodiment of a process 200 for comparing a new nucleotide or protein sequence with a database of sequences in order to determine the homology levels between the new sequence and the sequences in the database. The database of sequences can be a private database stored within the computer system 100, or a public database such as GENBANK, PIR OR SWISSPROT that is available through the Internet.

The process 200 begins at a start state 201 and then moves to a state 202 wherein the new sequence to be compared is stored to a memory in a computer system 100. As discussed above, the memory could be any type of memory, including RAM or an internal storage device.

The process 200 then moves to a state 204 wherein a database of sequences is opened for analysis and comparison. The process 200 then moves to a state 206 wherein the first sequence stored in the database is read into a memory on the computer. A comparison is then performed at a state 210 to determine if the first sequence is the same as the second sequence. It is important to

note that this step is not limited to performing an exact comparison between the new sequence and the first sequence in the database. Well-known methods are known to those of skill in the art for comparing two nucleotide or protein sequences, even if they are not identical. For example, gaps can be introduced into one sequence in order to raise the homology level between the two tested sequences. The parameters that control whether gaps or other features are introduced into a  
5 sequence during comparison are normally entered by the user of the computer system.

Once a comparison of the two sequences has been performed at the state 210, a determination is made at a decision state 210 whether the two sequences are the same. Of course, the term "same" is not limited to sequences that are absolutely identical. Sequences that are within  
10 the homology parameters entered by the user will be marked as "same" in the process 200.

If a determination is made that the two sequences are the same, the process 200 moves to a state 214 wherein the name of the sequence from the database is displayed to the user. This state notifies the user that the sequence with the displayed name fulfills the homology constraints that were entered. Once the name of the stored sequence is displayed to the user, the process 200  
15 moves to a decision state 218 wherein a determination is made whether more sequences exist in the database. If no more sequences exist in the database, then the process 200 terminates at an end state 220. However, if more sequences do exist in the database, then the process 200 moves to a state 224 wherein a pointer is moved to the next sequence in the database so that it can be compared to the new sequence. In this manner, the new sequence is aligned and compared with  
20 every sequence in the database.

It should be noted that if a determination had been made at the decision state 212 that the sequences were not homologous, then the process 200 would move immediately to the decision state 218 in order to determine if any other sequences were available in the database for comparison.

25 Accordingly, one aspect of the present invention is a computer system comprising a processor, a data storage device having stored thereon a nucleic acid code of SEQ ID NOs. 1 to 17, 24, 25, 28, 74 and 79 to 132 or a polypeptide code of SEQ ID Nos 18 to 23, a data storage device having retrievably stored thereon reference nucleotide sequences or polypeptide  
30 sequences to be compared to the nucleic acid code of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 or polypeptide code of SEQ ID Nos. 18 to 23 and a sequence comparer for conducting the comparison. The sequence comparer may indicate a homology level between the sequences compared or identify structural motifs in the above described nucleic acid code of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 and polypeptide codes of SEQ ID Nos. 18 to 23, or it may identify structural motifs in sequences which are compared to these nucleic acid codes and  
35 polypeptide codes. In some embodiments, the data storage device may have stored thereon the

sequences of at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 or polypeptide codes of SEQ ID Nos. 18 to 23.

Another aspect of the present invention is a method for determining the level of homology between a nucleic acid code of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 and a reference nucleotide sequence, comprising the steps of reading the nucleic acid code and the reference nucleotide sequence through the use of a computer program which determines homology levels and determining homology between the nucleic acid code and the reference nucleotide sequence with the computer program. The computer program may be any of a number of computer programs for determining homology levels, including those specifically enumerated herein, including BLAST2N with the default parameters or with any modified parameters. The method may be implemented using the computer systems described above. The method may also be performed by reading 2, 5, 10, 15, 20, 25, 30, or 50 of the above described nucleic acid codes of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 through use of the computer program and determining homology between the nucleic acid codes and reference nucleotide sequences.

Figure 14 is a flow diagram illustrating one embodiment of a process 250 in a computer for determining whether two sequences are homologous. The process 250 begins at a start state 252 and then moves to a state 254 wherein a first sequence to be compared is stored to a memory. The second sequence to be compared is then stored to a memory at a state 256. The process 250 then moves to a state 260 wherein the first character in the first sequence is read and then to a state 262 wherein the first character of the second sequence is read. It should be understood that if the sequence is a nucleotide sequence, then the character would normally be either A, T, C, G or U. If the sequence is a protein sequence, then it should be in the single letter amino acid code so that the first and sequence sequences can be easily compared.

A determination is then made at a decision state 264 whether the two characters are the same. If they are the same, then the process 250 moves to a state 268 wherein the next characters in the first and second sequences are read. A determination is then made whether the next characters are the same. If they are, then the process 250 continues this loop until two characters are not the same. If a determination is made that the next two characters are not the same, the process 250 moves to a decision state 274 to determine whether there are any more characters either sequence to read.

If there aren't any more characters to read, then the process 250 moves to a state 276 wherein the level of homology between the first and second sequences is displayed to the user. The level of homology is determined by calculating the proportion of characters between the sequences that were the same out of the total number of sequences in the first sequence. Thus, if every character in a first 100 nucleotide sequence aligned with a every character in a second



sequence, the homology level would be 100%.

Alternatively, the computer program may be a computer program which compares the nucleotide sequences of the nucleic acid codes of the present invention, to reference nucleotide sequences in order to determine whether the nucleic acid code of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 differs from a reference nucleic acid sequence at one or more positions.

Optionally such a program records the length and identity of inserted, deleted or substituted nucleotides with respect to the sequence of either the reference polynucleotide or the nucleic acid code of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132. In one embodiment, the computer program may be a program which determines whether the nucleotide sequences of the nucleic acid codes of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 contain a biallelic marker or single nucleotide polymorphism (SNP) with respect to a reference nucleotide sequence. This single nucleotide polymorphism may comprise a single base substitution, insertion, or deletion, while this biallelic marker may comprise about one to ten consecutive bases substituted, inserted or deleted.

Another aspect of the present invention is a method for determining the level of homology between a polypeptide code of SEQ ID Nos. 18 to 23 and a reference polypeptide sequence, comprising the steps of reading the polypeptide code of SEQ ID Nos. 18 to 23 and the reference polypeptide sequence through use of a computer program which determines homology levels and determining homology between the polypeptide code and the reference polypeptide sequence using the computer program.

Accordingly, another aspect of the present invention is a method for determining whether a nucleic acid code of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 differs at one or more nucleotides from a reference nucleotide sequence comprising the steps of reading the nucleic acid code and the reference nucleotide sequence through use of a computer program which identifies differences between nucleic acid sequences and identifying differences between the nucleic acid code and the reference nucleotide sequence with the computer program. In some embodiments, the computer program is a program which identifies single nucleotide polymorphisms. The method may be implemented by the computer systems described above and the method illustrated in Figure 14. The method may also be performed by reading at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 and the reference nucleotide sequences through the use of the computer program and identifying differences between the nucleic acid codes and the reference nucleotide sequences with the computer program.

In other embodiments the computer based system may further comprise an identifier for identifying features within the nucleotide sequences of the nucleic acid codes of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 or the amino acid sequences of the polypeptide codes of SEQ ID

Nos. 18 to 23.

An "identifier" refers to one or more programs which identifies certain features within the above-described nucleotide sequences of the nucleic acid codes of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 or the amino acid sequences of the polypeptide codes of SEQ ID Nos. 18 to 23. In one embodiment, the identifier may comprise a program which identifies an open reading frame in the cDNAs codes of SEQ ID Nos 2 to 17 and 36 to 40.

Figure 15 is a flow diagram illustrating one embodiment of an identifier process 300 for detecting the presence of a feature in a sequence. The process 300 begins at a start state 302 and then moves to a state 304 wherein a first sequence that is to be checked for features is stored to a memory 115 in the computer system 100. The process 300 then moves to a state 306 wherein a database of sequence features is opened. Such a database would include a list of each feature's attributes along with the name of the feature. For example, a feature name could be "Initiation Codon" and the attribute would be "ATG". Another example would be the feature name "TAATAA Box" and the feature attribute would be "TAATAA". An example of such a database is produced by the University of Wisconsin Genetics Computer Group ([www.gcg.com](http://www.gcg.com)).

Once the database of features is opened at the state 306, the process 300 moves to a state 308 wherein the first feature is read from the database. A comparison of the attribute of the first feature with the first sequence is then made at a state 310. A determination is then made at a decision state 316 whether the attribute of the feature was found in the first sequence. If the attribute was found, then the process 300 moves to a state 318 wherein the name of the found feature is displayed to the user.

The process 300 then moves to a decision state 320 wherein a determination is made whether more features exist in the database. If no more features do exist, then the process 300 terminates at an end state 324. However, if more features do exist in the database, then the process 300 reads the next sequence feature at a state 326 and loops back to the state 310 wherein the attribute of the next feature is compared against the first sequence.

It should be noted, that if the feature attribute is not found in the first sequence at the decision state 316, the process 300 moves directly to the decision state 320 in order to determine if any more features exist in the database.

In another embodiment, the identifier may comprise a molecular modeling program which determines the 3-dimensional structure of the polypeptides codes of SEQ ID Nos. 18 to 23. In some embodiments, the molecular modeling program identifies target sequences that are most compatible with profiles representing the structural environments of the residues in known three-dimensional protein structures. (See, e.g., Eisenberg et al., U.S. Patent No. 5,436,850

issued July 25, 1995). In another technique, the known three-dimensional structures of proteins in a given family are superimposed to define the structurally conserved regions in that family. This protein modeling technique also uses the known three-dimensional structure of a homologous protein to approximate the structure of the polypeptide codes of SEQ ID Nos. 4 to 8. (See e.g., Srinivasan, et al., U.S. Patent No. 5,557,535 issued September 17, 1996). Conventional homology modeling techniques have been used routinely to build models of proteases and antibodies. (Sowdhamini et al., Protein Engineering 10:207, 215 (1997)). Comparative approaches can also be used to develop three-dimensional protein models when the protein of interest has poor sequence identity to template proteins. In some cases, proteins fold into similar three-dimensional structures despite having very weak sequence identities. For example, the three-dimensional structures of a number of helical cytokines fold in similar three-dimensional topology in spite of weak sequence homology.

The recent development of threading methods now enables the identification of likely folding patterns in a number of situations where the structural relatedness between target and template(s) is not detectable at the sequence level. Hybrid methods, in which fold recognition is performed using Multiple Sequence Threading (MST), structural equivalencies are deduced from the threading output using a distance geometry program DRAGON to construct a low resolution model, and a full-atom representation is constructed using a molecular modeling package such as QUANTA.

According to this 3-step approach, candidate templates are first identified by using the novel fold recognition algorithm MST, which is capable of performing simultaneous threading of multiple aligned sequences onto one or more 3-D structures. In a second step, the structural equivalencies obtained from the MST output are converted into interresidue distance restraints and fed into the distance geometry program DRAGON, together with auxiliary information obtained from secondary structure predictions. The program combines the restraints in an unbiased manner and rapidly generates a large number of low resolution model confirmations. In a third step, these low resolution model confirmations are converted into full-atom models and subjected to energy minimization using the molecular modeling package QUANTA. (See e.g., Aszódi et al., Proteins: Structure, Function, and Genetics, Supplement 1:38-42 (1997)).

The results of the molecular modeling analysis may then be used in rational drug design techniques to identify agents which modulate the activity of the polypeptide codes of SEQ ID Nos. 18 to 23.

Accordingly, another aspect of the present invention is a method of identifying a feature within the nucleic acid codes of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 or the polypeptide codes of SEQ ID Nos. 18 to 23 comprising reading the nucleic acid code(s) or the

polypeptide code(s) through the use of a computer program which identifies features therein and identifying features within the nucleic acid code(s) or polypeptide code(s) with the computer program. In one embodiment, computer program comprises a computer program which identifies open reading frames. In a further embodiment, the computer program identifies structural motifs in a polypeptide sequence. In another embodiment, the computer program comprises a molecular modeling program. The method may be performed by reading a single sequence or at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 or the polypeptide codes of SEQ ID Nos. 18 to 23 through the use of the computer program and identifying features within the nucleic acid codes or polypeptide codes with the computer program.

The nucleic acid codes of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 or the polypeptide codes of SEQ ID Nos. 18 to 23 may be stored and manipulated in a variety of data processor programs in a variety of formats. For example, the nucleic acid codes of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 or the polypeptide codes of SEQ ID Nos. 18 to 23 may be stored as text in a word processing file, such as MicrosoftWORD or WORDPERFECT or as an ASCII file in a variety of database programs familiar to those of skill in the art, such as DB2, SYBASE, or ORACLE. In addition, many computer programs and databases may be used as sequence comparers, identifiers, or sources of reference nucleotide or polypeptide sequences to be compared to the nucleic acid codes of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 or the polypeptide codes of SEQ ID Nos. 18 to 23. The following list is intended not to limit the invention but to provide guidance to programs and databases which are useful with the nucleic acid codes of SEQ ID Nos. 1 to 17, 24, 25, 28, 74 and 79 to 132 or the polypeptide codes of SEQ ID Nos. 18 to 23. The programs and databases which may be used include, but are not limited to: MacPattern (EMBL), DiscoveryBase (Molecular Applications Group), GeneMine (Molecular Applications Group), Look (Molecular Applications Group), MacLook (Molecular Applications Group), BLAST and BLAST2 (NCBI), BLASTN and BLASTX (Altschul et al, *J. Mol. Biol.* 215: 403 (1990)), FASTA (Pearson and Lipman, *Proc. Natl. Acad. Sci. USA*, 85: 2444 (1988)), FASTDB (Brutlag et al. *Comp. App. Biosci.* 6:237-245, 1990), Catalyst (Molecular Simulations Inc.), Catalyst/SHAPE (Molecular Simulations Inc.), Cerius<sup>2</sup>.DBAccess (Molecular Simulations Inc.), HypoGen (Molecular Simulations Inc.), Insight II, (Molecular Simulations Inc.), Discover (Molecular Simulations Inc.), CHARMm (Molecular Simulations Inc.), Felix (Molecular Simulations Inc.), DelPhi, (Molecular Simulations Inc.), QuanteMM, (Molecular Simulations Inc.), Homology (Molecular Simulations Inc.), Modeler (Molecular Simulations Inc.), ISIS (Molecular Simulations Inc.), Quanta/Protein Design (Molecular Simulations Inc.), WebLab (Molecular Simulations Inc.), WebLab Diversity Explorer (Molecular Simulations Inc.), Gene Explorer

(Molecular Simulations Inc.), SeqFold (Molecular Simulations Inc.), the EMBL/Swissprotein database, the MDL Available Chemicals Directory database, the MDL Drug Data Report data base, the Comprehensive Medicinal Chemistry database, Derwent's World Drug Index database, the BioByteMasterFile database, the Genbank database, and the Genseqn database. Many other  
5 programs and data bases would be apparent to one of skill in the art given the present disclosure.

Motifs which may be detected using the above programs include sequences encoding leucine zippers, helix-turn-helix motifs, glycosylation sites, ubiquitination sites, alpha helices, and beta sheets, signal sequences encoding signal peptides which direct the secretion of the encoded proteins, sequences implicated in transcription regulation such as homeoboxes, acidic  
10 stretches, enzymatic active sites, substrate binding sites, and enzymatic cleavage sites.

Throughout this application, various publications, patents, and published patent applications are cited. The disclosures of the publications, patents, and published patent specifications referenced in this application are all hereby incorporated by reference in their entireties into the present disclosure to more fully describe the state of the art to which this  
15 invention pertains.

### **EXAMPLES**

Several of the methods of the present invention are described in the following examples, which are offered by way of illustration and not by way of limitation. Many other  
20 modifications and variations of the invention as herein set forth can be made without departing from the spirit and scope thereof and therefore only such limitations should be imposed as are indicated by the appended claims.

#### **Example 1**

##### **Identification Of Biallelic Markers - DNA Extraction**

25 Donors were unrelated and healthy. They presented a sufficient diversity for being representative of a heterogeneous population. The DNA from 100 individuals was extracted and tested for the detection of the biallelic markers.

30 30 ml of peripheral venous blood were taken from each donor in the presence of EDTA. Cells (pellet) were collected after centrifugation for 10 minutes at 2000 rpm. Red cells were lysed by a lysis solution (50 ml final volume: 10 mM Tris pH7.6; 5 mM MgCl<sub>2</sub>; 10 mM NaCl). The solution was centrifuged (10 minutes, 2000 rpm) as many times as necessary to eliminate the residual red cells present in the supernatant, after resuspension of the pellet in the lysis solution.

35 The pellet of white cells was lysed overnight at 42°C with 3.7 ml of lysis solution

composed of:

- 3 ml TE 10-2 (Tris-HCl 10 mM, EDTA 2 mM) / NaCl 0.4 M
- 200  $\mu$ l SDS 10%
- 500  $\mu$ l K-proteinase (2 mg K-proteinase in TE 10-2 / NaCl 0.4 M).

5

For the extraction of proteins, 1 ml saturated NaCl (6M) (1/3.5 v/v) was added. After vigorous agitation, the solution was centrifuged for 20 minutes at 10000 rpm.

10

For the precipitation of DNA, 2 to 3 volumes of 100% ethanol were added to the previous supernatant, and the solution was centrifuged for 30 minutes at 2000 rpm. The DNA solution was rinsed three times with 70% ethanol to eliminate salts, and centrifuged for 20 minutes at 2000 rpm. The pellet was dried at 37°C, and resuspended in 1 ml TE 10-1 or 1 ml water. The DNA concentration was evaluated by measuring the OD at 260 nm (1 unit OD = 50  $\mu$ g/ml DNA). To determine the presence of proteins in the DNA solution, the OD 260 / OD 280 ratio was determined. Only DNA preparations having a OD 260 / OD 280 ratio between 1.8 and 2 were used in the subsequent examples described below.

15

The pool was constituted by mixing equivalent quantities of DNA from each individual.

#### Example 2

##### Identification Of Biallelic Markers: Amplification Of Genomic DNA By PCR

20

The amplification of specific genomic sequences of the DNA samples of Example 1 was carried out on the pool of DNA obtained previously. In addition, 50 individual samples were similarly amplified.

PCR assays were performed using the following protocol:

25

Final volume	25 $\mu$ l
DNA	2 ng/ $\mu$ l
MgCl <sub>2</sub>	2 mM
dNTP (each)	200 $\mu$ M
primer (each)	2.9 ng/ $\mu$ l
Ampli Taq Gold DNA polymerase	0.05 unit/ $\mu$ l
PCR buffer (10x = 0.1 M TrisHCl pH8.3 0.5M KCl)	1x

30

35

Each pair of first primers was designed using the sequence information of genomic DNA sequences of SEQ ID Nos 1 to 78 disclosed herein and the OSP software (Hillier & Green, 1991). This first pair of primers was about 20 nucleotides in length and had the sequences disclosed in Table 4 in the columns labeled "Position range of amplification primer"

in SEQ ID No.” and “Complementary position range of amplification primer in SEQ ID No.”.

Table 4

Amplicon	SEQ ID No	Primer name	Position range of amplification primer in SEQ ID		Primer name	Complementary position range of amplification primer in SEQ ID	
99-27943	1	B1	7938	7958	C1	8446	8465
99-27935	1	B2	21365	21385	C2	21845	21864
8-128	1	B3	65453	65471	C3	65856	65874
99-31960	1	B4	95034	95053	C4	95543	95563
99-24656	1	B5	107022	107040	C5	107495	107513
99-24634	1	B6	160770	160787	C6	161240	161257
99-24639	1	B7	160279	160298	C7	160785	160802
99-7652	1	B8	168813	168830	C8	169331	169351
99-16100	1	B9	170666	170686	C9	171153	171173
99-5862	1	B10	173065	173085	C10	173495	173514
99-5919	1	B11	189753	189771	C11	190187	190207
99-24658	1	B12	197116	197135	C12	197555	197572
8-303	1	B13	200587	200606	C13	200994	201012
8-300	1	B14	202440	202459	C14	202852	202871
8-299	1	B15	203061	203080	C15	203486	203505
8-296	1	B16	203943	203962	C16	204345	204362
99-24644	1	B17	204741	204758	C17	205222	205240
8-252	1	B18	204373	204392	C18	204777	204794
8-295	1	B19	205034	205052	C19	205435	205453
8-293	1	B20	205768	205786	C20	206175	206193
8-292	1	B21	206316	206333	C21	206725	206742
8-251	1	B22	207186	207205	C22	207616	207634
8-289	1	B23	208186	208205	C23	208588	208606
8-287	1	B24	208786	208805	C24	209189	209208
8-285	1	B25	209520	209539	C25	209931	209949
8-283	1	B26	210219	210237	C26	210620	210638
8-282	1	B27	210801	210820	C27	211205	211223
8-281	1	B28	211182	211199	C28	211596	211613

8-279	1	B29	212290	212309	C29	212698	212716
8-278	1	B30	212662	212679	C30	213091	213109
8-277	1	B31	213783	213803	C31	214062	214082
8-124	1	B32	214564	214581	C32	214965	214983
8-273	1	B33	215310	215328	C33	215711	215729
8-126	1	B34	216687	216704	C34	217082	217101
8-130	1	B35	216683	216702	C35	217091	217109
99-5897	1	B36	219408	219425	C36	219882	219899
99-24649	1	B37	220505	220522	C37	221004	221021
8-247	1	B38	223011	223028	C38	223549	223566
8-246	1	B39	223962	223978	C39	224588	224605
8-245	1	B40	224564	224581	C40	225044	225061
8-119	1	B41	230210	230227	C41	230622	230641
99-15870	1	B42	231278	231298	C42	231729	231747
99-16087	1	B43	235081	235101	C43	235515	235533
8-155	1	B44	238607	238626	C44	239029	239046
99-16038	1	B45	239405	239425	C45	239862	239880
8-136	1	B46	239606	239624	C46	240012	240029
99-15880	1	B47	241700	241717	C47	242151	242171
8-271	1	B48	243484	243502	C48	243909	243926
8-267	1	B49	245105	245123	C49	245518	245535
8-266	1	B50	245372	245389	C50	245775	245793
8-265	1	B51	245754	245772	C51	246155	246173
99-25940	1	B52	247513	247533	C52	248023	248043
99-16032	1	B53	248204	248223	C53	248588	248606
99-16105	1	B54	255697	255715	C54	256133	256152
99-16033	1	B55	259885	259902	C55	260324	260342
99-15875	1	B56	279626	279644	C56	280154	280173
99-16052	1	B57	300153	300170	C57	300660	300680
99-16047	1	B58	311615	311632	C58	312126	312144
99-25993	1	B59	315649	315668	C59	316129	316147



Amplicon	SEQ ID No	Primer name	Position range of amplification primer in SEQ ID		Primer name	Complementary position range of amplification primer in SEQ ID	
8-102	24	B60	1	20	C60	408	427
8-104	25	B61	1	21	C61	295	310
8-94	26	B62	1	18	C62	400	418
8-95	27	B63	1	18	C63	400	419
8-97	28	B64	1	18	C64	331	350
8-98	29	B65	1	19	C65	415	434
99-14021	30	B66	1	18	C66	465	477
99-14364	31	B67	64	82	C67	434	454
99-15056	32	B68	1	18	C68	482	502
99-15229	33	B69	1	20	C69	476	494
99-15232	34	B70	1	18	C70	467	485
99-15253	35	B71	1	19	C71	459	477
99-15256	36	B72	1	18	C72	439	456
99-15258	37	B73	1	19	C73	521	541
99-15261	38	B74	1	19	C74	481	501
99-15280	39	B75	1	18	C75	521	541
99-15355	40	B76	1	18	C76	471	489
99-15663	41	B77	1	18	C77	430	450
99-15664	42	B78	1	19	C78	483	502
99-15665	43	B79	1	20	C79	458	476
99-15668	44	B80	1	18	C80	439	459
99-15672	45	B81	1	18	C81	533	547
99-15682	46	B82	1	19	C82	482	500
99-16081	47	B83	1	18	C83	443	462
99-16082	48	B84	1	18	C84	512	532
99-20933	49	B85	1	19	C85	432	451
99-20977	50	B86	1	18	C86	491	511
99-20978	51	B87	1	19	C87	446	466
99-20981	52	B88	1	18	C88	429	449
99-20983	53	B89	1	19	C89	431	449
99-22310	54	B90	1	19	C90	446	466

99-25922	55	B91	1	21	C91	451	471
99-25950	56	B92	1	19	C92	479	499
99-25961	57	B93	1	20	C93	464	484
99-25965	58	B94	1	21	C94	451	470
99-25966	59	B95	1	21	C95	503	523
99-25967	60	B96	1	20	C96	473	493
99-25969	61	B97	1	21	C97	510	530
99-25972	62	B98	1	21	C98	428	448
99-25974	63	B99	1	21	C99	524	544
99-25977	64	B100	1	21	C100	520	540
99-25979	65	B101	1	19	C101	518	538
99-25985	66	B102	1	21	C102	449	469
99-25989	67	B103	1	19	C103	533	553
99-26146	68	B104	1	19	C104	436	456
99-26150	69	B105	1	19	C105	437	454
99-26153	70	B106	1	19	C106	429	449
99-26156	71	B107	1	18	C107	491	511
99-27345	72	B108	1	18	C108	531	551
99-30329	73	B109	1	18	C109	496	514
99-31798	74	B110	1	20	C110	491	511
99-5873	75	B111	11	28	C111	466	474
99-6012	76	B112	11	29	C112	477	490
99-6080	77	B113	11	28	C113	519	529
99-7308	78	B114	1	18	C114	469	489

Preferably, the primers contained a common oligonucleotide tail upstream of the specific bases targeted for amplification which was useful for sequencing.

5 Primers from the column labeled "Position range of amplification primer in SEQ ID No." contain the following additional PU 5' sequence: TGTAACGACGGCCAGT (SEQ ID No. 133); primers from the column labeled "Complementary position range of amplification primer in SEQ ID No." contain the following RP 5' sequence: CAGGAAACAGCTATGACC (SEQ ID No. 134).

10 The synthesis of these primers was performed following the phosphoramidite method, on a GENSET UFPS 24.1 synthesizer.

DNA amplification was performed on a Genius II thermocycler. After heating at 95°C for 10 min, 40 cycles were performed. Each cycle comprised: 30 sec at 95°C, 54°C for 1 min, and 30 sec at 72°C. For final elongation, 10 min at 72°C ended the amplification. The quantities of the amplification products obtained were determined on 96-well microtiter plates, using a fluorometer and Picogreen as intercalant agent (Molecular Probes).

### Example 3

#### Identification of Polymorphisms

##### a) Identification of Biallelic Markers from Amplified Genomic DNA of Example 2

The sequencing of the amplified DNA obtained in Example 2 was carried out on ABI 377 sequencers. The sequences of the amplification products were determined using automated dideoxy terminator sequencing reactions with a dye terminator cycle sequencing protocol. The products of the sequencing reactions were run on sequencing gels and the sequences were determined using gel image analysis (ABI Prism DNA Sequencing Analysis software (2.1.2 version)).

The sequence data were further evaluated to detect the presence of biallelic markers within the amplified fragments. The polymorphism search was based on the presence of superimposed peaks in the electrophoresis pattern resulting from different bases occurring at the same position as described previously.

The localization of the biallelic markers detected in the fragments of amplification are as shown below in Table 5.

Table 5

Biallelic Markers

Amplicon	BM	Marker Name	Polymorphism		SEQ ID No.	BM position in SEQID	Position of probes in SEQ ID No.		Probe No.
			Al1	Al2					
99-27943	A1	99-27943-150	G	C	1	8316	8304	8328	P1
99-27935	A2	99-27935-193	G	C	1	21672	21660	21684	P2
8-128	A3	8-128-33	C	T	1	65485	65473	65497	P3
99-31960	A4	99-31960-363	A	G	1	95396	95384	95408	P4
99-24656	A5	99-24656-260	A	G	1	107281	107269	107293	P5
99-24634	A6	99-24639-163	A	C	1	160640	160628	160652	P6
99-24639	A7	99-24634-108	A	T	1	160876	160864	160888	P7
99-7652	A8	99-7652-162	A	G	1	168974	168962	168986	P8
99-16100	A9	99-16100-147	A	G	1	170810	170798	170822	P9
99-5862	A10	99-5862-167	A	G	1	173358	173346	173370	P10

99-5919	A11	99-5919-215	A	G	1	189957	189945	189969	P11
99-24658	A12	99-24658-410	A	G	1	197163	197151	197175	P12
8-303	A13	8-303-235	A	G	1	200778	200766	200790	P13
8-300	A14	8-300-221	A	G	1	202651	202639	202663	P14
8-300	A15	8-300-193	A	G	1	202679	202667	202691	P15
8-299	A16	8-299-128	A	T	1	203378	203366	203390	P16
8-296	A17	8-296-213	A	T	1	204138	204126	204150	P17
8-252	A18	8-252-190	C	T	1	204605	204593	204617	P18
99-24644	A19	99-24644-194	A	G	1	204934	204922	204946	P19
8-295	A20	8-295-248	A	C	1	205206	205194	205218	P20
8-295	A21	8-295-125	C	T	1	205329	205317	205341	P21
8-293	A22	8-293-130	A	G	1	206064	206052	206076	P22
8-292	A23	8-292-198	A	G	1	206545	206533	206557	P23
8-251	A24	8-251-322	A	G	1	207313	207301	207325	P24
8-289	A25	8-289-322	A	G	1	208285	208273	208297	P25
8-287	A26	8-287-249	C	T	1	208960	208948	208972	P26
8-287	A27	8-287-86	A	T	1	209123	209111	209135	P27
8-285	A28	8-285-319	A	G	1	209631	209619	209643	P28
8-283	A29	8-283-278	G	C	1	210361	210349	210373	P29
8-283	A30	8-283-176	A	G	1	210463	210451	210475	P30
8-283	A31	8-283-153	G	C	1	210486	210474	210498	P31
8-283	A32	8-283-56	C	T	1	210583	210571	210595	P32
8-282	A33	8-282-345	G	C	1	210879	210867	210891	P33
8-282	A34	8-282-260	G	T	1	210964	210952	210976	P34
8-282	A35	8-282-245	A	C	1	210979	210967	210991	P35
8-282	A36	8-282-174	Variable motif		1	211050			P36
8-282	A37	8-282-92	A	T	1	211132	211120	211144	P37
8-281	A38	8-281-367	A	G	1	211247	211235	211259	P38
8-281	A39	8-281-299	A	G	1	211315	211303	211327	P39
8-281	A40	8-281-248	G	C	1	211366	211354	211378	P40
8-279	A41	8-279-197	A	C	1	212520	212508	212532	P41
8-278	A42	8-278-289	C	T	1	212821	212809	212833	P42
8-277	A43	8-277-250	C	T	1	213833	213821	213845	P43

8-124	A44	8-124-294	A	G	1	214857	214845	214869	P44
8-273	A45	8-273-383	A	G	1	215347	215335	215359	P45
8-130/ 8-126	A46	8-130-143 / 8-126-286	A	G	1	216967	216955	216979	P46
99-5897	A47	99-5897-143	A	C	1	219540	219528	219552	P47
99-24649	A48	99-24649-186	A	G	1	220836	220824	220848	P48
8-247	A49	8-247-337	A	G	1	223230	223218	223242	P49
8-246	A50	8-246-393	A	T	1	224213	224201	224225	P50
8-245	A51	8-245-270	A	G	1	224792	224780	224804	P51
8-119	A52	8-119-247	A	G	1	230395	230383	230407	P52
99-15870	A53	99-15870-400	A	G	1	231677	231665	231689	P53
99-16087	A54	99-16087-219	G	C	1	235315	235303	235327	P54
8-155	A55	8-155-258	C	T	1	238789	238777	238801	P55
99-16038	A56	99-16038-118	C	T	1	239763	239751	239775	P56
8-136	A57	8-136-267	C	T	1	239763	239751	239775	P57
99-15880	A58	99-15880-162	A	G	1	241861	241849	241873	P58
8-271	A59	8-271-227	A	G	1	243700	243688	243712	P59
8-267	A60	8-267-273	Deletion		1	245263			P60
8-266	A61	8-266-131	A	C	1	245663	245651	245675	P61
8-265	A62	8-265-208	A	G	1	245966	245954	245978	P62
8-265	A63	8-265-203	G	C	1	245971	245959	245983	P63
99-25940	A64	99-25940-182	C	T	1	247864	247852	247876	P64
99-16032	A65	99-16032-292	G	T	1	248315	248303	248327	P65
99-16105	A66	99-16105-152	A	G	1	255848	255836	255860	P66
99-16033	A67	99-16033-244	A	G	1	260099	260087	260111	P67
99-15875	A68	99-15875-165	C	T	1	279789	279777	279801	P68
99-16052	A69	99-16052-214	A	G	1	300365	300353	300377	P69
99-16047	A70	99-16047-115	A	G	1	312030	312018	312042	P70
99-25993	A71	99-25993-367	A	G	1	316014	316002	316026	P71
Amplicon	BM	Marker Name	Polymorphism		SEQ ID No.	BM position in SEQID	Position of probes in SEQ ID No.		Probe No.
			All1	All2					
8-102	A72	8-102-231	C	T	24	231	219	243	P72
8-104	A73	8-104-105	A	G	25	105	93	117	P73

8-94	A74	8-94-252	A	G	26	251	239	263	P74
8-95	A75	8-95-43	A	G	27	43	31	55	P75
8-97	A76	8-97-181	A	G	28	181	169	193	P76
8-97	A77	8-97-98	C	T	28	98	86	110	P77
8-98	A78	8-98-68	A	G	29	68	56	80	P78
99-14021	A79	99-14021-108	A	G	30	108	96	120	P79
99-14364	A80	99-14364-415	C	T	31	415	403	427	P80
99-15056	A81	99-15056-99	C	T	32	99	87	111	P81
99-15229	A82	99-15229-412	A	G	33	412	400	424	P82
99-15232	A83	99-15232-291	G	T	34	291	279	303	P83
99-15253	A84	99-15253-382	C	T	35	382	370	394	P84
99-15256	A85	99-15256-392	C	T	36	392	380	404	P85
99-15258	A86	99-15258-337	G	T	37	337	325	349	P86
99-15261	A87	99-15261-202	A	G	38	200	188	212	P87
99-15280	A88	99-15280-432	C	T	39	432	420	444	P88
99-15355	A89	99-15355-150	C	T	40	150	138	162	P89
99-15663	A90	99-15663-298	C	T	41	298	286	310	P90
99-15664	A91	99-15664-185	G	T	42	185	173	197	P91
99-15665	A92	99-15665-398	A	G	43	398	386	410	P92
99-15668	A93	99-15668-139	C	T	44	139	127	151	P93
99-15672	A94	99-15672-166	C	T	45	166	154	178	P94
99-15682	A95	99-15682-318	A	T	46	318	306	330	P95
99-16081	A96	99-16081-217	C	T	47	217	205	229	P96
99-16082	A97	99-16082-218	A	G	48	218	206	230	P97
99-20933	A98	99-20933-81	A	C	49	81	69	93	P98
99-20977	A99	99-20977-72	A	C	50	71	59	83	P99
99-20978	A100	99-20978-89	G	C	51	89	77	101	P100
99-20981	A101	99-20981-300	A	G	52	300	288	312	P101
99-20983	A102	99-20983-48	A	G	53	48	36	60	P102
99-22310	A103	99-22310-148	C	T	54	148	136	160	P103
99-25922	A104	99-25922-147	G	T	55	147	135	159	P104
99-25950	A105	99-25950-121	G	C	56	121	109	133	P105
99-25961	A106	99-25961-376	A	C	57	373	361	385	P106
99-25965	A107	99-25965-399	A	G	58	399	387	411	P107

99-25966	A108	99-25966-241	A	G	59	241	229	253	P108
99-25967	A109	99-25967-57	A	G	60	56	44	68	P109
99-25969	A110	99-25969-200	G	T	61	200	188	212	P110
99-25972	A111	99-25972-317	C	T	62	315	303	327	P111
99-25974	A112	99-25974-143	A	G	63	143	131	155	P112
99-25977	A113	99-25977-311	A	G	64	311	299	323	P113
99-25979	A114	99-25979-93	A	G	65	93	81	105	P114
99-25985	A115	99-25985-194	C	T	66	194	182	206	P115
99-25989	A116	99-25989-398	A	G	67	398	386	410	P116
99-26146	A117	99-26146-264	G	T	68	264	252	276	P117
99-26150	A118	99-26150-276	A	G	69	276	264	288	P118
99-26153	A119	99-26153-44	A	C	70	44	32	56	P119
99-26156	A120	99-26156-290	A	C	71	290	278	302	P120
99-27345	A121	99-27345-189	G	C	72	189	177	201	P121
99-30329	A122	99-30329-380	C	T	73	380	368	392	P122
99-31798	A123	99-31798-344	Variable motif		74	342			P123
99-5873	A124	99-5873-159	C	T	75	159	147	171	P124
99-6012	A125	99-6012-220	G	T	76	220	208	232	P125
99-6080	A126	99-6080-99	C	T	77	99	87	111	P126
99-7308	A127	99-7308-157	C	T	78	156	144	168	P127

BM refers to "biallelic marker". A111 and A112 refer respectively to allele 1 and allele 2 of the biallelic marker.

Certain biallelic markers of the invention are insertions or deletions, as indicated above.

- 5 In particular, the biallelic marker 8-282-174 comprises a variable motif AAAGG or GAAGGAAGGAAGGAAGGAAGA. The biallelic marker 8-267-273 comprises an insertion or deletion of GTGGGC. The biallelic marker 99-31798-344 comprises a variable motif CAT or ACTGTCTCTACCTCA

10

#### Example 4

##### Validation Of The Polymorphisms Through Microsequencing

The biallelic markers identified in Example 3 were further confirmed and their respective frequencies were determined through microsequencing. Microsequencing was

carried out for each individual DNA sample described in Example 1.

Amplification from genomic DNA of individuals was performed by PCR as described above for the detection of the biallelic markers with the same set of PCR primers (Table 3).

The preferred primers used in microsequencing were about 19 nucleotides in length and hybridized just upstream of the considered polymorphic base. According to the invention, the primers used in microsequencing are detailed in Table 6.

**Table 6**

Marker Name	Biallelic Marker	SEQ ID No.	Mis. 1	Position range of microsequencing primer mis. 1 in SEQ ID No.		Mis. 2	Complementary position range of microsequencing primer mis. 2 in SEQ ID No.	
99-27943-150	A1	1	D1	8297	8315	E1	8317	8335
99-27935-193	A2	1	D2	21653	21671	E2	21673	21691
8-128-33	A3	1	D3	65466	65484	E3	65486	65504
99-31960-363	A4	1	D4	95377	95395	E4	95397	95415
99-24656-260	A5	1	D5	107262	107280	E5	107282	107300
99-24639-163	A6	1	D6	160621	160639	E6	160641	160659
99-24634-108	A7	1	D7	160857	160875	E7	160877	160895
99-7652-162	A8	1	D8	168955	168973	E8	168975	168993
99-16100-147	A9	1	D9	170791	170809	E9	170811	170829
99-5862-167	A10	1	D10	173339	173357	E10	173359	173377
99-5919-215	A11	1	D11	189938	189956	E11	189958	189976
99-24658-410	A12	1	D12	197144	197162	E12	197164	197182
8-303-235	A13	1	D13	200759	200777	E13	200779	200797
8-300-221	A14	1	D14	202632	202650	E14	202652	202670
8-300-193	A15	1	D15	202660	202678	E15	202680	202698
8-299-128	A16	1	D16	203359	203377	E16	203379	203397
8-296-213	A17	1	D17	204119	204137	E17	204139	204157
8-252-190	A18	1	D18	204586	204604	E18	204606	204624
99-24644-194	A19	1	D19	204915	204933	E19	204935	204953
8-295-248	A20	1	D20	205187	205205	E20	205207	205225
8-295-125	A21	1	D21	205310	205328	E21	205330	205348
8-293-130	A22	1	D22	206045	206063	E22	206065	206083
8-292-198	A23	1	D23	206526	206544	E23	206546	206564
8-251-322	A24	1	D24	207294	207312	E24	207314	207332



8-289-322	A25	1	D25	208266	208284	E25	208286	208304
8-287-249	A26	1	D26	208941	208959	E26	208961	208979
8-287-86	A27	1	D27	209104	209122	E27	209124	209142
8-285-319	A28	1	D28	209612	209630	E28	209632	209650
8-283-278	A29	1	D29	210342	210360	E29	210362	210380
8-283-176	A30	1	D30	210444	210462	E30	210464	210482
8-283-153	A31	1	D31	210467	210485	E31	210487	210505
8-283-56	A32	1	D32	210564	210582	E32	210584	210602
8-282-345	A33	1	D33	210860	210878	E33	210880	210898
8-282-260	A34	1	D34	210945	210963	E34	210965	210983
8-282-245	A35	1	D35	210960	210978	E35	210980	210998
8-282-174	A36	1	D36			E36		
8-282-92	A37	1	D37	211113	211131	E37	211133	211151
8-281-367	A38	1	D38	211228	211246	E38	211248	211266
8-281-299	A39	1	D39	211296	211314	E39	211316	211334
8-281-248	A40	1	D40	211347	211365	E40	211367	211385
8-279-197	A41	1	D41	212501	212519	E41	212521	212539
8-278-289	A42	1	D42	212802	212820	E42	212822	212840
8-277-250	A43	1	D43	213814	213832	E43	213834	213852
8-124-294	A44	1	D44	214838	214856	E44	214858	214876
8-273-383	A45	1	D45	215328	215346	E45	215348	215366
8-130-143 / 8-126-286	A46	1	D46	216948	216966	E46	216968	216986
99-5897-143	A47	1	D47	219521	219539	E47	219541	219559
99-24649-186	A48	1	D48	220817	220835	E48	220837	220855
8-247-337	A49	1	D49	223211	223229	E49	223231	223249
8-246-393	A50	1	D50	224194	224212	E50	224214	224232
8-245-270	A51	1	D51	224773	224791	E51	224793	224811
8-119-247	A52	1	D52	230376	230394	E52	230396	230414
99-15870-400	A53	1	D53	231658	231676	E53	231678	231696
99-16087-219	A54	1	D54	235296	235314	E54	235316	235334
8-155-258	A55	1	D55	238770	238788	E55	238790	238808
99-16038-118	A56	1	D56	239744	239762	E56	239764	239782
8-136-267	A57	1	D57	239744	239762	E57	239764	239782

99-15880-162	A58	1	D58	241842	241860	E58	241862	241880
8-271-227	A59	1	D59	243681	243699	E59	243701	243719
8-267-273	A60	1	D60			E60		
8-266-131	A61	1	D61	245644	245662	E61	245664	245682
8-265-208	A62	1	D62	245947	245965	E62	245967	245985
8-265-203	A63	1	D63	245952	245970	E63	245972	245990
99-25940-182	A64	1	D64	247845	247863	E64	247865	247883
99-16032-292	A65	1	D65	248296	248314	E65	248316	248334
99-16105-152	A66	1	D66	255829	255847	E66	255849	255867
99-16033-244	A67	1	D67	260080	260098	E67	260100	260118
99-15875-165	A68	1	D68	279770	279788	E68	279790	279808
99-16052-214	A69	1	D69	300346	300364	E69	300366	300384
99-16047-115	A70	1	D70	312011	312029	E70	312031	312049
99-25993-367	A71	1	D71	315995	316013	E71	316015	316033
Marker Name	Biallelic Marker	SEQ ID No.	Mis. 1	Position range of microsequencing primer mis. 1 in SEQ ID No.		Mis. 2	Complementary position range of microsequencing primer mis. 2 in SEQ ID No.	
8-102-231	A72	24	D72	212	230	E72	232	250
8-104-105	A73	25	D73	86	104	E73	106	124
8-94-252	A74	26	D74	232	250	E74	252	270
8-95-43	A75	27	D75	24	42	E75	44	62
8-97-181	A76	28	D76	162	180	E76	182	200
8-97-98	A77	28	D77	79	97	E77	99	117
8-98-68	A78	29	D78	49	67	E78	69	87
99-14021-108	A79	30	D79	89	107	E79	109	127
99-14364-415	A80	31	D80	396	414	E80	416	434
99-15056-99	A81	32	D81	80	98	E81	100	118
99-15229-412	A82	33	D82	393	411	E82	413	431
99-15232-291	A83	34	D83	272	290	E83	292	310
99-15253-382	A84	35	D84	363	381	E84	383	401
99-15256-392	A85	36	D85	373	391	E85	393	411
99-15258-337	A86	37	D86	318	336	E86	338	356
99-15261-202	A87	38	D87	181	199	E87	201	219

99-15280-432	A88	39	D88	413	431	E88	433	451
99-15355-150	A89	40	D89	131	149	E89	151	169
99-15663-298	A90	41	D90	279	297	E90	299	317
99-15664-185	A91	42	D91	166	184	E91	186	204
99-15665-398	A92	43	D92	379	397	E92	399	417
99-15668-139	A93	44	D93	120	138	E93	140	158
99-15672-166	A94	45	D94	147	165	E94	167	185
99-15682-318	A95	46	D95	299	317	E95	319	337
99-16081-217	A96	47	D96	198	216	E96	218	236
99-16082-218	A97	48	D97	199	217	E97	219	237
99-20933-81	A98	49	D98	62	80	E98	82	100
99-20977-72	A99	50	D99	52	70	E99	72	90
99-20978-89	A100	51	D100	70	88	E100	90	108
99-20981-300	A101	52	D101	281	299	E101	301	319
99-20983-48	A102	53	D102	29	47	E102	49	67
99-22310-148	A103	54	D103	129	147	E103	149	167
99-25922-147	A104	55	D104	128	146	E104	148	166
99-25950-121	A105	56	D105	102	120	E105	122	140
99-25961-376	A106	57	D106	354	372	E106	374	392
99-25965-399	A107	58	D107	380	398	E107	400	418
99-25966-241	A108	59	D108	222	240	E108	242	260
99-25967-57	A109	60	D109	37	55	E109	57	75
99-25969-200	A110	61	D110	181	199	E110	201	219
99-25972-317	A111	62	D111	296	314	E111	316	334
99-25974-143	A112	63	D112	124	142	E112	144	162
99-25977-311	A113	64	D113	292	310	E113	312	330
99-25979-93	A114	65	D114	74	92	E114	94	112
99-25985-194	A115	66	D115	175	193	E115	195	213
99-25989-398	A116	67	D116	379	397	E116	399	417
99-26146-264	A117	68	D117	245	263	E117	265	283
99-26150-276	A118	69	D118	257	275	E118	277	295
99-26153-44	A119	70	D119	25	43	E119	45	63
99-26156-290	A120	71	D120	271	289	E120	291	309
99-27345-189	A121	72	D121	170	188	E121	190	208

99-30329-380	A122	73	D122	361	379	E122	381	399
99-31798-344	A123	74	D123			E123		
99-5873-159	A124	75	D124	140	158	E124	160	178
99-6012-220	A125	76	D125	201	219	E125	221	239
99-6080-99	A126	77	D126	80	98	E126	100	118
99-7308-157	A127	78	D127	137	155	E127	157	175

Mis 1 and Mis 2 respectively refer to microsequencing primers which hybridized with the coding strand or with the non-coding strand of the nucleotide sequences of the invention.

The microsequencing reaction was performed as follows :

5 After purification of the amplification products, the microsequencing reaction mixture was prepared by adding, in a 20 µl final volume: 10 pmol microsequencing oligonucleotide, 1 U Thermosequenase (Amersham E79000G), 1.25 µl Thermosequenase buffer (260 mM Tris HCl pH 9.5, 65 mM MgCl<sub>2</sub>), and the two appropriate fluorescent ddNTPs (Perkin Elmer, Dye Terminator Set 401095) complementary to the nucleotides at the polymorphic site of each

10 biallelic marker tested, following the manufacturer's recommendations. After 4 minutes at 94°C, 20 PCR cycles of 15 sec at 55°C, 5 sec at 72°C, and 10 sec at 94°C were carried out in a Tetrad PTC-225 thermocycler (MJ Research). The unincorporated dye terminators were then removed by ethanol precipitation. Samples were finally resuspended in formamide-EDTA

15 sequencing gel. The data were collected by an ABI PRISM 377 DNA sequencer and processed using the GENESCAN software (Perkin Elmer).

Following gel analysis, data were automatically processed with software that allows the determination of the alleles of biallelic markers present in each amplified fragment.

20 The software evaluates such factors as whether the intensities of the signals resulting from the above microsequencing procedures are weak, normal, or saturated, or whether the signals are ambiguous. In addition, the software identifies significant peaks (according to shape and height criteria). Among the significant peaks, peaks corresponding to the targeted site are identified based on their position. When two significant peaks are detected for the same position, each sample is categorized classification as homozygous or heterozygous type based

25 on the height ratio.

#### Example 5a

#### Association Study Between Schizophrenia And The Biallelic Markers Of The Invention

**Collection Of DNA Samples From Affected And Non-Affected Individuals****A) Affected population**

5 All the samples were collected from a large epidemiological study of schizophrenia undertaken in hospital centers of Quebec from October 1995 to April 1997. The population was composed of French Caucasian individuals. The study design consisted in the ascertainment of cases and two of their first degree relatives (parents or siblings).

As a whole, 956 schizophrenic cases were ascertained according to the following inclusion criteria:

- the diagnosis had been done by a psychiatrist;
- 10 - the diagnosis had been done at least 3 years before recruitment time, in order to exclude individuals suffering from transient manic-depressive psychosis or depressive disorders;
- the patient ancestors had been living in Quebec for at least 6 generations;
- it was possible to get a blood sample from 2 close relatives.

15 Among the 956 schizophrenic ascertained cases, 834 individuals were included in the study for the following reasons:

- for the included individual cases, the diagnosis of schizophrenia was established according to the DSM-IV (Diagnostic and Statistical Manual, Fourth edition, Revised 1994, American Psychiatric Press);
- 20 - samples from individuals suffering from schizoaffective disorder were discarded;
- individuals suffering from catatonic schizophrenia were also excluded from the population of schizophrenic cases;
- were also excluded the individuals having a first degree relative or 2 or more second degree relatives suffering from depression or mood disorder;
- 25 - individuals having had severe head trauma, severe obstretical complications, encephalitis, or meningitis before onset of symptoms were also excluded;
- has also been excluded from the population of schizophrenic cases a patient suffering from epilepsy and treated with anticonvulsants.

The age at onset was not added as an inclusion criteria.

30 **B) Unaffected population**

Control cases were respectively ascertained based on the following cumulative criteria:

- the individual must not be affected by schizophrenia or any other psychiatric disorder;
- the individual must have 35 years old or more;
- the individual must belong to the French-Canadian population;
- 35 - the individual must have one or two first degree relative available for blood sampling.

Controls were matched with cases sex when possible.

### **C) Cases and Control Populations Selected for the Association Study**

The unaffected population retained for the study was composed of 241 individuals. The initial sample of the clinical study was composed of 215 cases and 214 controls. The controls were composed of 116 males and 98 females while the cases were composed of 154 males and 64 females. For each control, two first degree relatives (father, mother, sisters and brothers) were available. In order to match the sex of cases and controls, the parents of female controls were substituted for the female controls where possible and where the parents were known to be unaffected by schizophrenia or other psychosis. The parents of 27 female controls were thus substituted for the respective females, resulting in a total control sample size of 241 individuals. The composition of the control sample is detailed below in Table 7.

**Table 7**

Description of control samples

<b>Probands</b>	<b>187</b>
Male	116
Female	71
<b>Parents of probands</b>	<b>54</b>
Fathers	27
Mothers	27
<b>Total</b>	<b>241</b>

The association data that are presented below were obtained on a population size detailed in Table 8 below, wherein the individuals have been randomly selected from the populations detailed above.

**Table 8**

Cases and Control Populations Selected for the Association Study			
sample type		Cases	Controls
sample size		215	241
Gender			
	Male	151	143
	Female	64	98
Familial history of psychosis (FH)*			
	positive (FH+)	82	0
	none (FH-)	133	241

* : close relatives (first or second degree)		
----------------------------------------------	--	--

Both case and control populations form two groups, each group consisting of unrelated individuals that do not share a known common ancestor. Additionally, the individuals of the control population were selected among those having no family history of schizophrenia or schizophrenic disorder.

#### Genotyping of affected and control individuals

##### A) Results from the genotyping

The general strategy to perform the association studies was to individually scan the DNA samples from all individuals in each of the populations described above in order to establish the allele frequencies of biallelic markers, and among them the biallelic markers of the invention, in the diploid genome of the tested individuals belonging to each of these populations.

Allelic frequencies of every biallelic marker in each population (cases and controls) were determined by performing microsequencing reactions on amplified fragments obtained by genomic PCR performed on the DNA samples from each individual. Genomic PCR and microsequencing were performed as detailed above in Examples 1 to 4 using the described PCR and microsequencing primers.

##### Single biallelic marker frequency analysis

For each allele of the biallelic markers included in this study, the difference between the allelic frequency in the unaffected population and in the population affected by schizophrenia was calculated and the absolute value of the difference was determined. The more the difference in allelic frequency for a particular biallelic marker or a particular set of biallelic markers, the more probable an association between the genomic region harboring this particular biallelic marker or set of biallelic markers and schizophrenia. Allelic frequencies were also useful to check that the markers used in the haplotype studies meet the Hardy-Weinberg proportions (random mating).

The allelic frequencies of biallelic markers in the chromosome 13q31-q33 region between the affected and the unaffected population, using the sample population described above, is set forth in Table 9.

**Table 9**

Allelic frequencies of markers in different sub-samples

marker	alleles	all sample			
		cases			controls
		all	HF+	HF-	

99-20978/89	C/G	0.51	0.47	0.51	0.55
99-20983/48	A/G	0.30	0.28	0.33	0.29
99-20981/300	A/G	0.54	0.51	0.55	0.56
99-20977/72	A/C	0.40	0.41	0.38	0.35
99-6080/99	C/T	0.58	0.57	0.57	0.55
99-15229/412	A/G	0.54	0.52	0.55	0.53
99-22310/148	C/T	0.46	0.48	0.44	0.47
99-15232/291	C/T	0.46	0.48	0.43	0.47
99-14021/108	A/G	0.46	0.48	0.44	0.47
8-98/68	A/G	0.20	0.18	0.23	0.19
8-97/98	C/T	0.78	0.75	0.81	0.80
99-6012/220	C/T	0.20	0.19	0.23	0.19
8-95/43	A/G	0.18	0.20	0.18	0.21
99-7308/157	C/T	0.39	0.42	0.36	0.39
99-14364/415	C/T	0.38	0.40	0.36	0.39
99-15672/166	C/T	0.51	0.47	0.54	0.54
99-15668/139	C/T	0.58	0.56	0.62	0.65
99-15665/398	A/G	0.72	0.67	0.72	0.76
99-15663/298	C/T	0.72	0.67	0.72	0.76
99-15664/185	C/T	0.69	0.62	0.72	0.72
99-15682/318	A/T	0.35	0.40	0.34	0.32
99-20933/81	A/C	0.43	0.41	0.42	0.40
99-16081/217	C/T	0.43	0.38	0.46	0.39
99-16082/218	A/G	0.33	0.31	0.35	0.32
99-5862/167	C/T	0.47	0.43	0.44	0.51
99-16100/147	A/G	0.48	0.44	0.45	0.50
99-7652/162	A/G	0.49	0.46	0.46	0.52
99-5919/215	A/G	0.66	0.71	0.69	0.60
99-5897/143	A/C	0.58	0.61	0.53	0.59
99-15870/400	A/G	0.32	0.38	0.27	0.33
99-16032/292	A/C	0.61	0.62	0.64	0.58
99-15880/162	A/G	0.62	0.63	0.65	0.58
99-16038/118	A/G	0.38	0.36	0.35	0.42
99-15875/165	C/T	0.58	0.57	0.57	0.63
99-16033/244	C/T	0.55	0.57	0.49	0.54
99-16047/115	C/T	0.73	0.75	0.68	0.73

In the association study described herein, several individual biallelic markers were



shown to be significantly associated with schizophrenia. In particular, several of the chromosome 13q31-q33 region biallelic markers (99-16038/118 (A56), 99-7652/162 (A8), 99-5919/215 (A11), 99-15875/165 (A68), 99-16032/292 (A9)) showed significant association with schizophrenia in both familial and sporadic schizophrenia cases. The significance of the absolute value of the difference of allelic frequency of the individual biallelic markers in the affected and the unaffected population is set forth in 2, with several biallelic marker having allelic frequency differences with p-values approaching or less than 0.05, biallelic marker 99-5919/215 (A11) having a p-value of less than 0.01. Figure 2 also shows the physical order of certain specific biallelic markers. These results show that several biallelic markers individually associated with schizophrenia are physically located in a particular region of significance, the subregion of the chromosome 13q31-q33 region referred to herein as Region D.

#### Haplotype frequency analysis

Analysis of markers Haplotype analysis for association of chromosome 13q31-q33-related biallelic markers and schizophrenia was performed by estimating the frequencies of all possible 2, 3 and 4 marker haplotypes in the affected and control populations described above. Haplotype estimations were performed by applying the Expectation-Maximization (EM) algorithm (Excoffier and Slatkin, 1995), using the EM-HAPLO program (Hawley et al., 1994) as described above. Estimated haplotype frequencies in the affected and control population were compared by means of a chi-square statistical test (one degree of freedom).

#### Haplotype association results in schizophrenia cases

The results of the haplotype analysis using the chromosome 13q31-q33-related biallelic markers biallelic markers is shown in Figure 3. In particular, the figures show the most significant haplotypes using the biallelic markers: 99-16047/115 (A10), 99-16033/244 (A67), 99-16038/118 (A56), 99-15875/165 (A68), 99-16032/292 (A9), 99-5897/143 (A47), 99-15880/162 (A58), 99-16082/218 (A97), 99-5919/215 (A11), 99-7652/162 (A8), 99-16100/147 (A9), 99-5862/167 (A10).

A number of biallelic marker haplotypes were shown to be significantly associated with schizophrenia. A first preferred haplotype (HAP287 of Figure 3) consisting of four biallelic markers (99-16038/118 (A56), 99-16082/218 (A97), (99-7652/162 (A8) and 99-16100/147 (A9)) is highly significantly associated with schizophrenia in both total cases and sporadic cases. Figure 4 shows the characteristics of this haplotype. This haplotype presented a p-value of  $3.1 \times 10^{-7}$  and an odd-ratio of 4.01 for total cases and a p-value of  $3.9 \times 10^{-6}$  and an odd-ratio of 3.88 for sporadic cases. Phenotypic permutation tests confirmed the statistical significance of these results. Estimated haplotype frequencies were 13.8% in total cases, 13.5% in the sporadic cases, and 3.8 % in the controls.

Several other significant haplotypes are listed in Figure 3, including several 2-, 3- and 4-marker haplotypes. Considered to be highly significantly associated with schizophrenia are the most significant 2-marker haplotype (HAP1 consisting of biallelic markers 99-15875/165 (A68) and 99-5919/215 (A11)) and the most significant 3-marker haplotype (HAP67 consisting of biallelic markers 99-16038/118 (A56), 99-16082/218 (A97) and 99-7652/162 (A8)).

Further preferred significant haplotypes considered associated with schizophrenia are haplotypes having p-values above a desired threshold level are also; all the haplotypes listed in Figure 3 present p-values below  $1.0 \times 10^{-2}$  for 2-marker haplotypes,  $1.0 \times 10^{-4}$  for 3-marker haplotypes, and  $1.0 \times 10^{-5}$  for 4-marker haplotypes. All of the biallelic markers presented in Figure 4 except for 1 (99-16047/115 (A70)) are involved in haplotypes having a p-value above these threshold levels. Figure 3 shows several 2-marker haplotypes, HAP1 to HAP8, having p-values ranging from  $1.0 \times 10^{-2}$  to  $1.2 \times 10^{-3}$ , several 3-marker haplotypes, HAP67 to HAP76, having p-values ranging from  $1.3 \times 10^{-5}$  to  $1.0 \times 10^{-4}$  and several 4-marker haplotypes, HAP287 to HAP291, having p-values ranging from  $8.2 \times 10^{-7}$  to  $3.1 \times 10^{-7}$ . Figure 4 shows biallelic markers involved in significant haplotypes having significance thresholds of  $1.0 \times 10^{-2}$ ,  $1.0 \times 10^{-4}$ , and  $1.0 \times 10^{-5}$  for 2-, 3- and 4-marker haplotypes, respectively.

Several 2-, 3- and 4-marker haplotypes, HAP 1, HAP8, HAP70, HAP71, HAP75, HAP76, HAP288, HAP290 and HAP291, often comprised the biallelic marker 99-5919/215 (A11) allele A. Furthermore, several 2-, 3- and 4-marker haplotypes, HAP7, HAP67, HAP69, HAP75, HAP287 AND HAP288, often comprised the biallelic marker 99-16038/118 (A56) allele G.

#### Example 5b

#### Association Study Between Schizophrenia And The Biallelic Markers Of The Invention

#### Collection Of DNA Samples From Affected And Non-Affected Individuals

Biallelic markers of the invention were further analyzed in the French Canadian population described above. For this analysis, the proband case population under study consisted of 139 individuals, the control population consisted of 141 individuals, as described in Table 10 below.

**Table 10**

Cases and Control Populations Selected for the Association Study			
Sample type		Cases	Controls
Sample size		139	141
Gender			
	Male	94	96

	Female	45	45
Familial history of psychosis (FH)*			
	positive (FH+)	76	0
	none (FH-)	63	141
* : close relatives (first or second degree)			

### Genotyping of affected and control individuals

#### A) Results from the genotyping

5 The general strategy for performing the association studies was to individually scan the DNA samples from all individuals in each of the populations described above in order to establish the allele frequencies of biallelic markers, and among them the biallelic markers of the invention, in the diploid genome of the tested individuals belonging to each of these populations.

10 Allelic frequencies of every biallelic marker in each population (cases and controls) were determined by performing microsequencing reactions on amplified fragments obtained by genomic PCR performed on the DNA samples from each individual. Genomic PCR and microsequencing were performed as detailed above in Examples 1 to 3 using the described PCR and microsequencing primers.

#### Single biallelic marker frequency analysis

15 For each allele of the biallelic markers included in this study, the difference between the allelic frequency in the unaffected population and in the population affected by schizophrenia was calculated and the absolute value of the difference was determined. The allelic frequencies of between the affected and the unaffected population in the regions is set forth in Table 11, using the sample population described above and in Table 10. The more the difference in allelic frequency for a particular biallelic marker or a particular set of biallelic markers, the more probable an association between the genomic region harboring this particular biallelic marker or set of biallelic markers and schizophrenia. Allelic frequencies were also useful to check that the markers used in the haplotype studies meet the Hardy-Weinberg proportions (random mating).

Table 11

25 Allelic frequencies of markers in different sub-samples (%)

Marker	polymorphism	Cases			All controls
		All cases	HF+	HF-	
99-20978/89	C/G	50,37	47,26	54,03	55,43
99-20983/48	A/G	30,37	28,67	32,5	26,52

99-20977/72	A/C	41,01	42,11	39,68	34,4
99-20981/300	A/G	52,17	51,33	53,17	60
99-6080/99	C/T	58,82	58	59,84	54,85
99-15229/412	A/G	54,92	52,86	57,26	51,88
99-22310/148	C/T	44,2	46,71	41,13	48,57
99-15232/291	G/T	43,85	46,43	40,83	49,28
99-14021/108	A/G	44,85	47,26	42,06	48,54
8-94/252	A/G	2,22	1,97	2,54	2,52
8-98/68	A/G	19,06	17,76	20,63	19,06
8-97/98	C/T	76,26	74,34	78,57	77,3
99-6012/220	G/T	20	18,49	21,77	18,79
99-7308/157	C/T	40,31	41,89	38,18	39,36
99-14364/415	C/T	39,93	40,79	38,89	40
8-95/43	A/G	20,29	20,39	20,16	22,14
99-15672/166	C/T	49,28	47,37	51,59	56,74
99-15668/139	C/T	58,21	56,16	60,66	66,67
99-15665/398	A/G	70,5	67,76	73,81	76,79
99-15663/298	C/T	70,5	67,76	73,81	76,95
99-15664/185	G/T	66,54	62,33	71,43	72,5
99-15682/318	A/T	35,27	39,58	29,82	32,66
99-20933/81	A/C	43,12	42,76	43,55	42,45
99-26146/264	G/T	39,62	38,67	40,91	38,85
99-25922/147	G/T	44,19	39,58	50	40,94
99-16081/217	C/T	42,28	38,82	46,67	36,74
99-16082/218	A/G	34,73	31,94	38,14	33,81
99-24656/260	A/G	48,87	49,32	48,31	54,04
99-24639/163	G/T	38,52	33,33	45	40,51
99-24634/108	A/T	44,85	42,67	47,54	50
99-7652/162	A/G	45,29	44,08	46,77	50,36
99-16100/147	A/G	44,66	42,75	46,77	48,89
99-5862/167	C/T	43,53	41,45	46,03	49,29
99-5919/215	A/G	69,42	71,05	67,46	60,28
99-24658/410	C/T	64,13	69,08	58,06	61,07
99-24644/194	A/G	39,42	41,22	37,3	40,51
99-5897/143	A/C	57,61	60,67	53,97	61,07
99-24649/186	C/T	67,75	67,33	68,25	62,95
99-15870/400	A/G	33,46	36,67	29,51	30,29
99-16038/118	A/G	34,53	36,18	32,54	43,62

99-15880/162	A/G	65,11	63,16	67,46	56,43
99-25940/182	A/G	59,42	56,67	62,7	52,59
99-16032/292	A/C	64,03	61,84	66,67	55,67
99-16033/244	C/T	54,51	56,76	51,69	56,44
99-15875/165	C/T	56,88	57,89	55,65	66,3
99-16047/115	C/T	71,69	74,67	68,03	75,19
99-25993/367	A/G	44,53	40,79	49,18	40,51
99-25989/398	A/G	32,81	33,33	32,2	27,86
99-25979/93	A/G	68,12	69,08	66,94	69,32
99-25969/200	G/T	36,67	38,67	34,17	38,85
99-25966/241	A/G	66,3	67,11	65,32	63,21
99-25961/376	A/C	39,63	42,57	36,07	37,31
99-25965/399	A/G	50,36	51,97	48,39	49,64
99-25977/311	A/G	72,01	67,76	77,59	73,72
99-25950/121	C/G	31,75	36	26,61	27,54
99-25974/143	A/G	25,55	28,29	22,13	22,7
99-26150/276	A/G	46,54	51,43	40,83	47,76
99-15258/337	G/T	25,55	26,97	23,77	24,1
99-15261/202	A/G	63,06	59,46	67,5	65,15
99-15256/392	C/T	64,96	61,33	69,35	65,3
99-15056/99	C/T	32,72	36,49	28,23	31,11
99-15280/432	C/T	42,28	44	40,16	38,97
99-15355/150	C/T	72,3	70,39	74,6	68,79
99-15253/382	C/T	63,04	62,67	63,49	62,95
99-5873/159	C/T	78,1	79,05	76,98	77,34

### Haplotype frequency analysis

Analysis of markers Haplotype analysis for association of chromosome 13q31-q33-related biallelic markers and schizophrenia was performed by estimating the frequencies of all possible 2, 3 and 4 marker haplotypes in the affected and control populations described above. Haplotype estimations were performed by applying the Expectation-Maximization (EM) algorithm (Excoffier and Slatkin, 1995), using the EM-HAPLO program (Hawley et al., 1994) as described above. Estimated haplotype frequencies in the affected and control population were compared by means of a chi-square statistical test (one degree of freedom).

### Haplotype association results in schizophrenia cases

Haplotype studies yielded significant results indicating an association of the nucleotide sequences of the invention with schizophrenia. Significant results are shown in Figures 5 and 6,

including descriptions of the frequency of the haplotype leading to the maximum chi square test (reference no. (1) in figures), the test of the frequency of a particular haplotype in cases vs in controls (reference no. (2) in figures) and the p-value assuming that the test has a chi-square distribution with 1 degree of freedom (ddl) (reference no. (3) in figures). The results of the

5 haplotype analysis using 28 preferred biallelic markers of the invention, 99-24656-260 (A5), 99-24639-163 (A6), 99-24634-108 (A7), 99-7652-162 (A8), 99-16100-147 (A9), 99-5862-167 (A10), 99-5919-215 (A11), 99-24658-410 (A12), 99-24644-194 (A19), 99-5897-143 (A47), 99-24649-186 (A48), 99-16038-118 (A56), 99-15880-162 (A58), 99-25940-182 (A64), 99-16032-292 (A9), 99-16033-244 (A67), 99-15875-165 (A68), 99-16047-115 (A70), 99-25950-121

10 (A105), 99-25961-376 (A106), 99-25965-399 (A103), 99-25966-241 (A108), 99-25969-200 (A110), 99-25974-143 (A112), 99-25977-311 (A113), 99-25979-93 (A114), 99-25989-398 (A116), and 99-26150-276 (A118) are shown in Figures 5 and 6. Figures 5 and 6 also show the physical order of the biallelic markers comprising the haplotypes.

Figure 5 shows the results of the haplotype analysis using the following biallelic

15 markers located on the approximately 319kb sequence of SEQ ID No. 1: 99-24656-260 (A5), 99-24639-163 (A6), 99-24634-108 (A7), 99-7652-162 (A8), 99-16100-147 (A9), 99-5862-167 (A10), 99-5919-215 (A11), 99-24658-410 (A12), 99-24644-194 (A19), 99-5897-143 (A47), 99-24649-186 (A48), 99-16038-118 (A56), 99-15880-162 (A58), 99-25940-182 (A64), 99-16032-292 (A65), 99-16033-244 (A67), 99-15875-165 (A68), and 99-16047-115 (A70).

Figure 6 shows the results of the haplotype analysis using the following biallelic

20 markers located on the approximately 319kb of SEQ ID No. 1 as well as additional biallelic markers located on the human chromosome 13q31-q33 locus: 99-16038-118 (A56), 99-15880-162 (A58), 99-25940-182 (A64), 99-16032-292 (A65), 99-16033-244 (A67), 99-15875-165 (A68), 99-16047-115 (A70), 99-25950-121 (A105), 99-25961-376 (A106), 99-25965-399

25 (A102), 99-25966-241 (A108), 99-25969-200 (A110), 99-25974-143 (A112), 99-25977-311 (A113), 99-25979-93 (A114), 99-25989-398 (A116), and 99-26150-276 (A119).

A number of biallelic marker haplotypes were shown to be significantly associated with schizophrenia.

Several preferred haplotype all showing highly significant association with

30 schizophrenia and including various 2-, 3- and 4- marker haplotypes are haplotypes 817, 818 and 819, 137, 138, 1 and 2 of Figure 6, and haplotypes 970, 154 and 1 of Figure 5. The p-values, odd-ratios and estimated haplotype frequencies are further described in Figures 5 and 6. In particular, the two marker haplotype 1 of Figure 5 consisting of biallelic markers 99-5862-167 (A10) and 99-15875-165 (A68) showed a highly significant p-value of  $7.8 \times 10^{-5}$  and an

35 odd-ratio of 1.61. Haplotype 818 of Figure 6 consisting of four biallelic markers (99-16032-292

(A65), 99-25969-200 (A110), 99-25977-311 (A113), and 99-25989-398 (A116)) presented a p-value of  $3.1 \times 10^{-7}$  and an odd-ratio of 9.08. Another example showing significance is haplotype 817 of Figure 6-consisting of four biallelic markers (99-16033-244 (A67), 99-15875-165 (A68), 99-25950-121 (A105) and 99-25979-93 (A114)), presented a p-value of  $2.4 \times 10^{-7}$  and an odd-ratio of 100. Phenotypic permutation tests confirmed the statistical significance of these results. Estimated haplotype frequencies were 10.5% in cases and 0 % in the controls. Haplotype 970 of Figure 5 consisting of four biallelic markers (99-5919-215 (A11), 99-24658-410 (A12), 99-15875-165 (A68), and 99-16047-115 (A10)) presented a p-value of  $7.8 \times 10^{-7}$  and an odd-ratio of 2.41. Phenotypic permutation tests confirmed the statistical significance of these results. Estimated haplotype frequencies were 25.7% in cases and 12.5 % in the controls.

Several other significant haplotypes are listed in Figures 5 and 6, including several 2-, 3- and 4-marker haplotypes. Considered to be highly significantly associated with schizophrenia are the most significant 2-marker haplotypes (for example haplotype 1 of Figure 5 noted above and the most significant 3-marker haplotypes (for example haplotype 137 of Figure 6 consisting of biallelic markers (99-15875-165 (A68), 99-16047-115 (A70) and 99-25950-121 (A105)).

Further preferred significant haplotypes considered associated with schizophrenia are haplotypes having p-values above a desired threshold level; all the haplotypes listed in Figures 5 and 6 present p-values below  $1.0 \times 10^{-2}$  for 2-marker haplotypes,  $1.0 \times 10^{-4}$  for 3-marker haplotypes, and  $1.0 \times 10^{-5}$  for 4-marker haplotypes. Figures 5 and 6 show several 2-marker haplotypes, haplotypes 1 to 9 and haplotypes 1 to 5 of Figures 5 and 6 respectively, having p-values ranging from  $7.8 \times 10^{-5}$  to  $8.6 \times 10^{-3}$ , several 3-marker haplotypes, haplotypes 154 to 163 and 137 to 141 of Figures 5 and 6 respectively, having p-values ranging from  $3.9 \times 10^{-6}$  to  $1.1 \times 10^{-4}$  and several 4-marker haplotypes, haplotypes 970 to 973 and 817 to 836 of Figures 5 and 6 respectively, having p-values ranging from  $2.4 \times 10^{-7}$  to  $7.3 \times 10^{-6}$ .

Additionally, a particularly large number of the significant 2-, 3- and 4-marker haplotypes often comprised the biallelic markers A65, A12, A67, A70, A107, A110, A116 and most commonly A68 (99-15875-165), allele T.

The statistical significance of the results obtained for the haplotype analysis was evaluated by a phenotypic permutation test reiterated 100 times on a computer. For this computer simulation, data from the affected and control individuals were pooled and randomly allocated to two groups which contained the same number of individuals as the case-control populations used to produce the data summarized in figures 5 and 6. A haplotype analysis was then run on these artificial groups for the markers included in the haplotypes showing strong

association with schizophrenia. This experiment was reiterated 100 times and the results are shown in the columns of Figures 5 and 6 labelled "Haplotype test by permutation procedure". For a given haplotype, these results demonstrate the number of obtained (simulated) haplotypes having a p-value comparable to the one obtained for the given haplotype among 100 iterations. These results, set forth in Figures 5 and 6 validate the statistical significance of the association between the haplotypes and schizophrenia.

#### Example 5c

#### Association Study Between Schizophrenia and the Biallelic Markers of the Invention in French Canadian Samples

##### Collection Of DNA Samples From Affected And Non-Affected Individuals

Biallelic markers of the present invention were further genotyped in French Canadian samples as described above in order to compare the association of the 1st and the 2nd portion of Region D with schizophrenia. The population used in the study was the same as described above with the exception that 2 male FH+ cases were not included.

The biallelic markers analyzed in the study include 34 preferred biallelic markers of the invention located in Region D of the chromosome 13q31-33 region. Included in the analysis were the 14 following biallelic markers from the first of two portions of Region D: 99-26150/276 (A118), 99-26156/290 (A120), 99-26153/44 (A119), 99-25985/194 (A115), 99-25974/143 (A112), 99-25977/311 (A114), 99-25972/317 (A111), 99-25965/399 (A107), 99-25961/376 (A106), 99-25966/241 (A108), 25967/57 (A109), 99-25969/200 (A110), 99-25979/93 (A114) and 99-25989/398 (A116). Included in the analysis were also the 20 following biallelic markers from the second of two portions of Region D: 99-25993/367 (A71), 99-16047/115 (A70), 99-15875/165 (A68), 99-16033/244 (A67), 99-16032/292 (A65), 99-25940/182 (A64), 99-15880/162 (A58), 99-16038/118 (A56), 99-15870/400 (A53), 99-24649/186 (A48), 99-5897/143 (A47), 99-24644/194 (A19), 99-24658/410 (A12), 99-5919/215 (A11), 99-5862/167 (A10), 99-16100/147 (A9), 99-7652/162 (A8), 99-24634/108 (A7), 99-24639/163 (A6) and 99-24656/260 (A5).

##### Single biallelic marker association results in schizophrenia cases

Single biallelic marker studies yielded significant results, indicating an association of the nucleotide sequences of the invention with schizophrenia. Biallelic markers used in the analysis included the set of 34 biallelic markers shown in Table 11 below, 14 biallelic markers of which were located on the first of two portions of Region D, and 20 of which were located on the second portion. The distribution of markers is shown in Table 12 below. As summarized in Table 13, analyses using these biallelic markers demonstrated a significant association with schizophrenia for 5 markers on the second portion of Region D.



Table 11

REGION	CONTIG	SNPS GENOTYPED	POLYMORPHISM	FREQUENCY IN CONTROLS
D	1 <sup>st</sup> portion	99-26150/276	A/G	50
		99-26156/290	A/C	69
		99-26153/44	A/C	61
		99-25985/194	C/T	29
		99-25974/143	A/G	25
		99-25977/311	A/G	73
		99-25972/317	C/T	72
		99-25965/399	A/G	49
		99-25961/376	A/C	40
		99-25966/241	A/G	63
		99-25967/57	A/G	43
		99-25969/200	G/T	40
		99-25979/93	A/G	72
		99-25989/398	A/G	29
	2 <sup>nd</sup> portion	99-25993/367	A/G	44
		99-16047/115	C/T	73
		99-15875/165	C/T	63
		99-16033/244	C/T	54
		99-16032/292	A/C	58
		99-25940/182	A/G	53
		99-15880/162	A/G	58
		99-16038/118	A/G	42
		99-15870/400	A/G	33
		99-24649/186	C/T	65
		99-5897/143	A/C	59
		99-24644/194	A/G	39
		99-24658/410	C/T	58
		99-5919/215	A/G	60
		99-5862/167	C/T	51
		99-16100/147	A/G	50
		99-7652/162	A/G	52
		99-24634/108	A/T	53
		99-24639/163	G/T	44
		99-24656/260	A/G	54

Table 12

Region	No. of Biallelic markers ( $\sigma$ )	Mean frequency ( $\sigma$ )	Mean inter-marker distance ( $\sigma$ )
D 1 <sup>st</sup> half	14 (14)	0.34 (0.07)	7 (6.3)
D 2 <sup>nd</sup> half	20 (8)	0.42 (0.06)	11 (13)
D 1 <sup>st</sup> and 2 <sup>nd</sup> half	34 (22)	0.39 (0.07)	10.3 (11)

Analysis of markers Haplotype analysis for association of chromosome 13q31-q33-related biallelic markers and schizophrenia was performed by estimating the frequencies of all possible 2, 3 and 4 marker haplotypes in the affected and control populations described above. Haplotype estimations were performed by applying the Expectation-Maximization (EM) algorithm (Excoffier and Slatkin, 1995), using the EM-HAPLO program (Hawley et al., 1994) as described above.

#### Haplotype association results in schizophrenia cases

Significant results were also obtained in haplotype studies indicating an association of the nucleotide sequences of the invention with schizophrenia.

The present inventors having previously demonstrated highly significant association of biallelic markers located on the Region D subregion of the human chromosome 13q31-q33 locus with disease. Using the Omnibus LR test which compares the profile of haplotype frequencies, and Haplo-maxM test which is based on haplotype differences for each haplotype in two groups, Figures 7 and 8 describe the results of an analysis of the first and second portions of Region D which demonstrated an association of the second portion of Region D with schizophrenia.

For combinations of 2 and 3 biallelic markers, one likelihood ratio test is obtained based on the haplotype frequency values calculated using the E-M algorithm. A permutation procedure was used, where data from the affected and control individuals was pooled and randomly allocated to two groups which contained the same number of individuals as the case-control populations used to produce the data. A haplotype analysis was then run on these artificial groups for the markers included in the haplotypes showing strong association with schizophrenia. This experiment was reiterated 100 times. For a given haplotype, these results demonstrate the number of obtained (simulated) haplotypes having a p-value comparable to the one obtained for the given haplotype among 100 iterations.

Figure 7 shows a comparison of the LR test value distributions of haplotype frequencies in the two portions of Region D. This association of the second portion of Region D with schizophrenia is shown using both 2-marker and 3-marker combinations. The distribution of LR test values in the different regions was analyzed using a Kruskal-Wallis rank test, a chi-square test with  $r-1$  degrees of freedom, where  $r$  represents the number of value sets compared. As shown, the significance of the association is demonstrated by a chi-square value (one degree of freedom) of 74.405 and a p-value of less than  $1 \times 10^{-10}$  for 2 marker combinations, and a chi-square value (one degree of freedom) of 228.72 and a p-value of  $1 \times 10^{-10}$  for 3- marker combinations.

Another association analysis approach based on haplotype frequency differences, referred to as the Haplo-maxM test, was conducted using region D biallelic markers. For one combination of markers having  $h$  haplotypes,  $h$  differences of haplotype frequencies can be compared via a Pearson chi-square statistic (one degree of freedom). The haplo-max test selects the difference showing the maximum positive test value between cases versus controls (rejecting test values based on rare haplotype frequencies, i.e., with an estimated number of haplotypes inferior to 10); for one combination of markers there is therefore one Max-M test value. The results of the Haplo-maxM test using Region D biallelic markers are shown in Figure 8.

Figure 8 shows the distribution of haplo-maxM test values obtained for both 2-marker and 3-marker combinations in the two portions of Region D, demonstrating an association of the second portion of Region D with schizophrenia. The comparison of the distribution of Haplo-maxM test values on the two regions was analyzed using a Kruskal-Wallis rank test, a chi-square test with  $r-1$  degrees of freedom, where  $r$  represents the number of value sets compared. As shown, the significance of the association is demonstrated by a chi-square value (one degree of freedom) of 34.839 and a p-value of less than  $3.58 \times 10^{-9}$  for 2 marker combinations, and a chi-square value (one degree of freedom) of 13.773 and a p-value of  $2.6 \times 10^{-4}$  for 3- marker combinations.

The results from the haplo-maxM tests further confirms the association shown using the Omnibus LR test results.

Results of association studies discussed above using biallelic markers of the invention are further summarized in Table 13 below, showing a significant association of the biallelic markers with schizophrenia in both single biallelic marker and haplotype analysis.

Table 13

	Single-point Analysis		Multi-point analysis (Haplotype-based analysis)		
	No. of allelic freq differences > 10%	No. Significant allelic tests	Omnibus LR TEST *		
			2-mks	3-mks	4-mks
Region D, 1st portion	0	0	0,03	0,05	0,06
Region D, 2nd portion	0	5	0,30	0,30	0,31

\* percentage of significant tests (5% level of significance)

Cases (N=213) / Controls (N=241)

#### Example 5d

#### Association Study Between Bipolar Disorder and the Biallelic Markers of the Invention

**Description of study design**

Biallelic markers of the invention were analyzed in bipolar disorder cases. As in examples above, single and multi-point analyses showed a significant association of the markers of the invention, of Region D of the chromosome 13q33 locus, and more particularly of a sub-region of Region D with bipolar disorder.

**A) Description of the Affected population**

All the samples were collected from a study of bipolar disorder undertaken in a hospital located south of Buenos Aires, Argentina, generally representing a population estimated at about 400,000 inhabitants. Patients were evaluated by four doctors in 1994 and 1995. The study design involved in the ascertainment of cases and their first degree relatives (parents or siblings). 514 individuals were available for the study. This group consisted of 158 subjects from 51 different families, and 356 independent subjects.

As a whole, bipolar disorder cases were ascertained according to the diagnosis of bipolar disorder established by the DSM-IV (Diagnostic and Statistical Manual, Fourth edition, Revised 1994, American Psychiatric Press);

Available for consideration for each coded case were also age, sex, nationality of parents and grand parents, ethnic origin, familial composition, marital state, socio-economic level, educational level, professional situation, employment, recreational activities, age of onset of psychiatric symptoms, age of first consultation, occurrences of obstetric or prenatal incidents, suicide attempts, other medical conditions, treatment for or occurrence of a neurological condition, familial occurrence of symptoms, previous or concurrent use of psychotropic drugs, other admissions to a hospital or medical treatments, and diagnostic reason for admission including (a) DSM-IV diagnosis and (b) symptoms first presented on admission to hospital.

Available for study were 226 bipolar disorder ascertained cases of which 203 were independent cases. This group consisted of 51 cases from 51 families, 20 cases in relatives thereof, and 155 independent cases. Upon elimination of 3 cases from the initial independent 155 cases due to discovery of a familial relation, the total number of independent cases was 203.

Cases were classified according to bipolar disorder type. The cases included 115 bipolar disorder type I individuals (including 1 rapid cycling case), 67 bipolar disorder type II individuals (including 1 rapid cycling case), 18 unclassified bipolar disorder cases, and 3 cases which remained unclassified due to lack of or inconsistent information.

The 203 independent cases were examined for a familial history of psychosis. 53 of these cases reported an occurrence of psychosis (characterized as schizophrenia or bipolar disorder) among first degree relatives (father, mother, brothers, sisters or children).

**B) Description of the Unaffected population**

Available for study were 201 controls which had not been affected by any psychiatric difficulties or reported any familial history of psychiatric difficulties. Available for consideration were also age, sex and ethnic origin of the unaffected population.

5

**C) Case and Control Populations Selected for the Association Study**

For the association study, the case population under study consisted of 201 individuals selected from the 226 total cases above; the control population consisted of 198 individuals selected from the 201 controls described above.

10

The association data that are presented in the Example 5d below were obtained on a population size detailed in Table 14 below.

**Table 14**

Cases and Control Populations Selected for the Association Study			
Sample type		Cases	Controls
Sample size		201	198
Gender			
	Male	68	81
	Female	124	117
	Missing	9	
Ethnic origin			
	Caucasian	182	177
	Non caucasian	5	21
	Missing	14	
Familial history of psychosis (FH)*			
	positive (FH+)	54	0
	none (FH-)	147	198
* : close relatives (first degree)			

Both case and control populations form two groups, each group consisting of unrelated individuals that do not share a known common ancestor.

15

**Genotyping of affected and control individuals**

The general strategy was to individually scan the DNA samples from all individuals in each of the populations described above in order to establish the allele frequencies of biallelic markers, and among them the biallelic markers of the invention, in the diploid genome of the tested individuals belonging to each of these populations.

20

Allelic frequencies of every biallelic marker in each population (cases and controls) were determined by performing microsequencing reactions on amplified fragments obtained by

genomic PCR performed on the DNA samples from each individual. Genomic PCR and microsequencing were performed as detailed above in Examples 1 to 4 using the described PCR and microsequencing primers.

#### Association analysis

5 The association analysis included 30 preferred biallelic markers of the invention located in Region D of the chromosome 13q31-33 region. Included in the analysis were the 14 following biallelic markers from the first of two subjective portions of Region D: 99-26150/276 (A118), 99-26156/290 (A120), 99-26153/44 (A119), 99-25985/194 (A115), 99-25974/143 (A112), 99-25977/311 (A113), 99-25972/317 (A111), 99-25965/399 (A107), 99-25961/376 (A106), 99-25966/241 (A108), 25967/57 (A109), 99-25969/200 (A110), 99-25979/93 (A114) and 99-25989/398 (A116). Included in the analysis were also the 16 following biallelic markers from the second of two portions of Region D: 99-25993/367 (A71), 99-16047/115 (A70), 99-15875/165 (A68), 99-16033/244 (A67), 99-16032/292 (A66), 99-25940/182 (A65), 99-15880/162 (A58), 99-16038/118 (A56), 99-15870/400 (A53), 99-24649/186 (A48), 99-5897/143 (A47), 99-24644/194 (A19), 99-5919/215 (A11), 99-5862/167 (A10), 99-16100/147 (A9), and 99-7652/162 (A8).

#### A) Single biallelic marker association results in bipolar disorder cases

For each allele of the biallelic markers included in this study, the difference between the allelic frequency in the unaffected population and in the population affected by bipolar disorder was calculated and the absolute value of the difference was determined. The set of biallelic markers and their allelic frequencies included in this study are set forth in Table 15. The more the difference in allelic frequency for a particular biallelic marker or a particular set of biallelic markers, the more probable an association between the genomic region harboring this particular biallelic marker or set of biallelic markers and bipolar disorder. Allelic frequencies were also useful to check that the markers used in the haplotype studies meet the Hardy-Weinberg proportions (under random mating assumptions)

Table 15

REGION	CONTIG	POSITION ON CONTIG	SNPS GENOTYPED	POLYMORPHISM	FREQUENCY IN CONTROLS
		168,02	99-26150/276	A/G	62,93
		173,29	99-26156/290	A/C	72,42
		177,01	99-26153/44	A/C	52,66
		186,41	99-25985/194	C/T	28,87
		190,15	99-25974/143	A/G	31,79
		216,43	99-25977/311	A/G	63,82

Region D  
first Half

224,62	99-25972/317	C/T	72,32
236,64	99-25965/399	A/G	58,24
244,82	99-25961/376	A/C	44,35
254,70	99-25966/241	A/G	66,18
257,85	99-25967/57	A/G	42,44
261,23	99-25969/200	G/T	35,76
263,67	99-25979/93	A/G	67,15
269,39	99-25989/398	A/G	35,88
299,02	99-25993/367	A/G	47,38
303,04	99-16047/115	C/T	69,01
335,02	99-15875/165	C/T	61,3
354,81	99-16033/244	C/T	50,3
366,51	99-16032/292	A/C	62,87
367,14	99-25940/182	A/G	54,39
372,98	99-15880/162	A/G	62,72
375,28	99-16038/118	A/G	37,29
383,41	99-15870/400	A/G	29,65
394,16	99-24649/186	C/T	66,57
395,27	99-5897/143	A/C	52,6
409,93	99-24644/194	A/G	38,29
424,95	99-5919/215	A/G	60,63
441,62	99-5862/167	C/T	46,53
444,00	99-16100/147	A/G	48,84
445,84	99-7652/162	A/G	49,7
<b>TOTAL</b>	<b>30</b>		

(1) : frequency (%) in caucasian controls (N=177) of the first allele (alphabetic order)

Region D was arbitrarily split in two halves (D 1<sup>st</sup> half and D 2<sup>nd</sup> half) for purpose of the analysis.

5

The present inventors have previously demonstrated significant association of biallelic markers located on the Region D subregion of the human chromosome 13q31-33 region with disease. Using a set of 30 biallelic markers shown in Table 15, D 1<sup>st</sup> half contained 14 markers and D 2<sup>nd</sup> half contained 16 markers.

10

Table 15 also shows the physical order of the biallelic markers on Region D of the human chromosome 13q31-q33 region. The mean intermarker distances of the biallelic markers on the first and the second subjective portions of Region D were as listed below in Table 16.

**Table 16**

Region	Mean Inter-marker distance (std)
D 1 <sup>st</sup> half	7.80(6.33)
D 2 <sup>nd</sup> half	9.79 (8.78)
D 1 <sup>st</sup> and 2 <sup>nd</sup> half	9.58 (8.46)

The analysis using selected Region D biallelic markers of the invention demonstrated a significant association with bipolar disorder for the second portion of Region D. The analysis was conducted using the sample population described above with 182 caucasian cases and 177 caucasian controls selected from the total case and control group.

One biallelic marker in particular, 99-15875/165(A68), located on the second half of Region D, demonstrated a significant association with disease at a significance level of better than 5% (corresponding to an absolute logarithm (p-value) of 1.3).

#### **B) Haplotype association results in bipolar disorder cases**

Haplotype analysis for association of chromosome 13q31-q33-related biallelic markers and bipolar disorder was performed by estimating the frequencies of all possible 2, 3 and 4 marker haplotypes in the affected and control populations described above. Haplotype frequencies estimations were performed by applying the Expectation-Maximization (EM) algorithm (Excoffier and Slatkin, 1995), modified by Nicholas Schork.

Significant results were obtained in haplotype studies indicating an association of the nucleotide sequences of the invention with bipolar disorder. The haplotype analysis as shown in the Figures 9A, 9B, 10A, 10B, 11A and 11B was conducted using the sample population described above, using 182 caucasian cases and 177 caucasian controls selected from the total case and control group.

Using the Omnibus LR test which compares the profile of haplotype frequencies, and Haplo-maxM test which is based on haplotype frequencies differences for each haplotype in two groups, Figures 9A, 9B, 10A, 10B, 11A and 11B show the results of a comparison of the first and second portions of Region D which demonstrated an association of the second portion of Region D with bipolar disorder.

#### **a - Omnibus LR tests values**

For a given combination of 2, 3 or 4 biallelic markers, one likelihood ratio test (LR test) is obtained based on the haplotype frequencies values calculated using the E-M algorithm.

Figures 9A and 9B show a comparison of the LR test value distributions of haplotype frequencies in the two portions of Region D. This association of the second portion of Region



D with bipolar disorder is shown using both 2-marker and 3-marker combinations. A Kruskal Wallis rank test was used to compare LR test values distributions in the two subjective portions of Region D. This test has an asymptotic Chi-square distribution, under the null hypothesis of no difference between the sets compared, with (r-1) degrees of freedom, where r represents the number of sets compared. Here, we compare the 2 portions of region D, so  $r=2$ , and the asymptotic Chi-square distribution has 1 degree of freedom. As shown, the significance of the association is demonstrated by a chi-square value (one degree of freedom) of 46.62 and a p-value of  $8.62 \times 10^{-12}$  for 2 marker combinations, and a chi-square value (one degree of freedom) of 124.72 and a p-value of  $5.86 \times 10^{-29}$  for 3- marker combinations.

#### 10      **b - Haplo-max tests values**

Another association analysis approach based on haplotype frequencies differences, referred to as the Haplo-max test, was conducted using region D biallelic markers. The haplo-max test selects the difference showing the maximum positive (maxM) or negative (maxS) test value between cases versus controls (rejecting test values based on rare haplotype frequencies, i.e, with an estimated number of haplotypes carriers inferior to 10) ; for one combination of markers there is therefore one Max-M and one Max-S test values.

Figures 10A and 10B show the distribution of haplo-maxM test values obtained for both 2-marker and 3-marker combinations in the two portions of Region D, demonstrating an association of the second portion of Region D with bipolar disorder. The comparison of the distribution of Haplo-maxM test values in the two regions was analyzed using a Kruskal-Wallis rank test, a chi-square test with 1 degree of freedom. As shown, the significance of the association is demonstrated by a chi-square value of 29.07 and a p-value  $6.98 \times 10^{-8}$  for 2 marker combinations, and a chi-square value of 98.63 and a p-value of  $3.04 \times 10^{-23}$  for 3- marker combinations.

Figures 11A and 11B show the distribution of Haplo-maxS test values again obtained for all 2-marker and 3-marker combinations in the two portions of Region D, demonstrating an association of the second portion of Region D with bipolar disorder. The comparison of the distributions of Haplo-maxS test values in the two portions was analyzed using a Kruskal-Wallis rank test with one degree of freedom. As shown, the significance of the association is demonstrated by a chi-square value of 34.6 and a p-value of  $4.05 \times 10^{-9}$  for 2 marker combinations, and a chi-square value of 98.31 and a p-value of  $3.58 \times 10^{-23}$  for 3- marker combinations.

The results from the haplo-maxM and haplo-maxS tests thus further confirm the

association shown using the Omnibus LR test results.

### Example 5e

#### Identification of Common Haplotypes in Three Schizophrenia Populations

5

Based on the results above indicating an association of biallelic markers located in Region D3 with schizophrenia, further biallelic markers in Region D3 were screened in three different populations of schizophrenia cases and controls. The D3 region is an approximately 150kb genomic region containing the g35030 gene.

10

Several significant haplotypes comprised of markers in or flanking the g35030 gene and common to the three populations were identified. These haplotypes demonstrated significance in Omnibus LR and individual haplotype frequency tests. Furthermore, the estimated frequencies of the haplotypes were highly similar for each of the different populations.

15

The samples used were obtained from a French Canadian population, an Eastern European population, and a United States sample. The French Canadian schizophrenia sample is described above in Example 5a, and Tables 7 and 8. The United States schizophrenia sample is described in Table 17 below. The Eastern European sample, is described in Table 18 below.

Table 17

United States Schizophrenia Cases and Control Populations (United States Caucasians)			
Sample type		Cases	Random US Controls
Sample size		131	188
Ethnic origin			
	European Caucasians (26 female, 66 male)	92	
	Ashkenazi caucasians (7 female, 17 female)	24	
	Other Caucasians (7 female, 8 male)	15	
Familial history of psychosis (FH)			
	positive (FH+)	133	
	none (FH-)	147	198

20

Table 18

Eastern European (Rogaev) Schizophrenia Cases and Control Populations			
Sample type		Cases	Controls
Sample size		183	183
Ethnic origin: European Causasians (97 female, 86 male)			
Familial history of psychosis (FH)			
	positive (FH+)	55	
	none (FH-)	128	

For each of the three populations, multipoint analyses were carried out using biallelic markers from the D3 region. The haplotypic frequency differences between case and controls was examined, reported as MaxM for the maximum positive difference, and MaxS as the maximum negative difference, and the profile of haplotype frequency differences, where the distribution of haplotype frequency differences is described using the Omnibus LR test.

Several significant two-marker haplotypes were found to be in common among the three populations. One haplotype, referred to haplotype 2, is described further in Tables 19 and 20. Other haplotypes are comprised of biallelic markers which are in linkage disequilibrium with one of the biallelic markers of haplotype 2, and thus also provide significant results. Biallelic markers 99-15875-165 (A68) (allele T) and 99-5862-167, involved in haplotype 2, demonstrated a range of estimation of haplotype frequencies of 23.6 to 24.5 in cases, and 15.5 to 17.6 in controls. Again, similar differences in frequency estimates were seen with other haplotypes. Markers 99-15875-165 (A68) (allele T) and 99-5862-167 are located in or flanking the g35030 gene.

Table 19 shows the results of analysis of estimated haplotype frequencies for haplotype 2, comprising biallelic markers 99-15875-165 (A68) (allele T) and 99-5862-167 (A10) (allele T). Presented are the results of a likelihood ratio test (LR test), demonstrating a significant p-value in each of the three populations; the results of the LR test on which a permutation procedure has been carried out (Omnibus), demonstrating a significant p-value of  $5 \times 10^{-3}$  for French Canadian samples,  $8 \times 10^{-2}$  for Eastern European (Rogaev) samples, and  $1 \times 10^{-2}$  for U.S. samples; and the estimated haplotype frequencies in cases and controls, and their positive or negative difference (MaxM). Notably, the haplotype frequencies in the three populations are highly similar at 23.60, 23.90 and 24.50 for cases and 15.5, 17.6 and 15.5 for controls.

Table 19

HAP2: 99-15875/165 (T), 99-5862/167 (T)							
POP	haplotype	Pvalue-hap	Omnibus (1000 perm)	freq cas	freq controls	Difference	
French Canadian	TT LD Normalized:-0.16)	2,20E-03	5,00E-03	23,60	15,5	8,10	axM
Eastern European		4,60E-02	8,70E-02	23,90	17,6	6,3	axM
U.S.		8,60E-03	1,30E-02	24,50	15,5	9	axM

5

**Example 6****Preparation of Antibody Compositions to the g35030 protein**

Substantially pure protein or polypeptide is isolated from transfected or transformed cells containing an expression vector encoding the g35030 protein or a portion thereof. The concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows:

10

**A. Monoclonal Antibody Production by Hybridoma Fusion**

15

Monoclonal antibody to epitopes in the g35030 protein or a portion thereof can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., Nature 256:495 (1975) or derivative methods thereof. Also see Harlow, E., and D. Lane. 1988. Antibodies A Laboratory Manual. Cold Spring Harbor Laboratory. pp. 53-242.

20

Briefly, a mouse is repetitively inoculated with a few micrograms of the g35030 protein or a portion thereof over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant

fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, (1980), and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. et al. Basic Methods in Molecular Biology Elsevier, New York. Section 21-2.

#### B. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogeneous epitopes in the g35030 protein or a portion thereof can be prepared by immunizing suitable non-human animal with the g35030 protein or a portion thereof, which can be unmodified or modified to enhance immunogenicity. A suitable non-human animal is preferably a non-human mammal is selected, usually a mouse, rat, rabbit, goat, or horse. Alternatively, a crude preparation which has been enriched for g35030 concentration can be used to generate antibodies. Such proteins, fragments or preparations are introduced into the non-human mammal in the presence of an appropriate adjuvant (e.g. aluminum hydroxide, RIBI, etc.) which is known in the art. In addition the protein, fragment or preparation can be pretreated with an agent which will increase antigenicity, such agents are known in the art and include, for example, methylated bovine serum albumin (mBSA), bovine serum albumin (BSA), Hepatitis B surface antigen, and keyhole limpet hemocyanin (KLH). Serum from the immunized animal is collected, treated and tested according to known procedures. If the serum contains polyclonal antibodies to undesired epitopes, the polyclonal antibodies can be purified by immunoaffinity chromatography.

Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. Techniques for producing and processing polyclonal antisera are known in the art, see for example, Mayer and Walker (1987). An effective immunization protocol for rabbits can be found in Vaitukaitis, J. et al. J. Clin. Endocrinol. Metab. 33:988-991 (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. et al., (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12  $\mu$ M). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: Manual of Clinical Immunology, 2d Ed. (Rose and Friedman, Eds.) Amer. Soc. For Microbiol., Washington, D.C. (1980).

Antibody preparations prepared according to either the monoclonal or the polyclonal protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

The disclosures of all issued patents, published PCT applications, scientific references or other publications cited herein are incorporated herein by reference in their entireties.

Although this invention has been described in terms of certain preferred embodiments, other embodiments which will be apparent to those of ordinary skill in the art of view of the disclosure herein are also within the scope of this invention. Accordingly, the scope of the invention is intended to be defined only by reference to the appended claims.

15

**Sequence listing free text**

The following free text appears in the accompanying Sequence Listing:

	schizophrenia
5	associated
	proteins
	biallelic
	markers
	regulatory
10	region
	gene
	exon
	complement
	polymorphic base
15	or
	allele
	deletion
	of
	basic
20	protease
	cleavage
	site
	probe
	primer
25	downstream
	upstream
	amplification
	sequencing
	oligonucleotide

The disclosures of the following references are incorporated herein by reference in their entireties:

- Abbondanzo S.J. et al. (1993) *Methods in Enzymology*, Academic Press, New York. pp. 803-823.
- 5 Ajioka R.S. et al. (1997) *Am. J. Hum. Genet.* 60:1439-1447.
- Altschul et al., 1990, *J. Mol. Biol.* 215(3):403-410;
- Altschul et al., 1993, *Nature Genetics* 3:266-272
- Altschul et al., 1997, *Nuc. Acids Res.* 25:3389-3402
- Ames, R.S. et al. *J. Immunol. Methods* 184:177-186(1995)
- 10 Anton M. et al., 1995, *J. Virol.*, 69: 4600-4606.
- Araki K et al. (1995) *Proc. Natl. Acad. Sci. US A.* 92(1):160-4.
- Ashkenazi, A. et al. *PNAS* 88:10535-10539(1991)
- Ausubel et al. (1989) *Current Protocols in Molecular Biology*, Green Publishing Associates and Wiley Interscience, N.Y.
- 15 Bartunek, P. et al. *Cytokine* 8(1):14-20 (1996)
- Baubonis W. (1993) *Nucleic Acids Res.* 21(9):2025-9.
- Beaucage et al., *Tetrahedron Lett* 1981, 22: 1859-1862
- Better, M. et al. *Science* 240:1041-1043(1988)
- Bradley A., (1987) *Production and analysis of chimaeric mice.* In: E.J. Robertson (Ed.),
- 20 *Teratocarcinomas and embryonic stem cells: A practical approach.* IRL Press, Oxford, pp. 113.
- Brinkman U. et al. *J. Immunol. Methods* 182:41-50(1995)
- Brown EL, Belagaje R, Ryan MJ, Khorana HG, *Methods Enzymol* 1979;68:109-151
- Brutlag et al. *Comp. App. Biosci.* 6:237-245, 1990
- Burton, D.R. et al. *Advances in Immunology* 57:191-280(1994)
- 25 Carlson, N.G. et al. *J. Biol. Chem.* 272(17):11295-11301(1997)
- Chai H. et al. (1993) *Biotechnol. Appl. Biochem.* 18:259-273.
- Chee et al. (1996) *Science.* 274:610-614.
- Chen and Kwok *Nucleic Acids Research* 25:347-353 1997
- Chen et al. (1987) *Mol. Cell. Biol.* 7:2745-2752.
- 30 Chen et al., 1987, *Mol. Cell. Biol.*, 7: 2745-2752.
- Chen, Z. et al. *Cancer Res.* 58(16): 3668-3678(1998)
- Chou J.Y. (1989) *Mol. Endocrinol.* 3:1511-1514.
- Clark A.G. (1990) *Mol. Biol. Evol.* 7:111-122.
- Coles et al. *Hum. Mol. Genet.*, 7:791-800, 1998
- 35 Compton J. (1991) *Nature.* 350(6313):91-92.



- Davis et al., Basic Methods in Molecular Biology, ed., Elsevier Press, NY, 1986
- Dempster et al., (1977) J.R. Stat. Soc., 39B:1-38.
- Deng, B. et al. Blood 92(6): 1981-1988(1998)
- Dent D.S. and Latchman D.S. (1993) The DNA mobility shift assay. In: Transcription Factors:  
5 A Practical Approach (Latchman DS, ed.) Oxford: IRL Press. pp1-26.
- Eckner R. et al. (1991) EMBO J. 10:3513-3522.
- Excoffier L. and Slatkin M. (1995) Mol. Biol. Evol., 12(5): 921-927.
- Feldman and Steg, 1996, Medecine Sciences, synthese, 12:47-55
- Fell, H.P. et al. J. Immunol. 146:2446-2452(1991)
- 10 Flotte et al. (1992) Am. J. Respir. Cell Mol. Biol. 7:349-356.
- Fodor et al. (1991) Science 251:767-777.
- Fraleigh et al. (1979) Proc. Natl. Acad. Sci. USA. 76:3348-3352.
- Fried M. and Crothers D.M. (1981) Nucleic Acids Res. 9:6505-6525.
- Fuller S. A. et al. (1996) Immunology in Current Protocols in Molecular Biology, Ausubel et  
15 al. Eds, John Wiley & Sons, Inc., USA.
- Furth P.A. et al. (1994) Proc. Natl. Acad. Sci. USA. 91:9302-9306.
- Garner M.M. and Revzin A. (1981) Nucleic Acids Res. 9:3047-3060.
- Geysen H. Mario et al. 1984. Proc. Natl. Acad. Sci. U.S.A. 81:3998-4002
- Ghosh and Bacchawat (1991) Targeting of liposomes to hepatocytes, IN: Liver Diseases,  
20 Targeted diagnosis and therapy using specific receptors and ligands. Wu et al. Eds.,  
Marcel Dekker, New York, pp. 87-104.
- Gillies, S.D. et al. J. Immunol. Methods 125:191-202(1989)
- Gillies, S.O. et al. PNAS 89:1428-1432(1992)
- Gonnet et al., 1992, Science 256:1443-1445;
- 25 Gopal (1985) Mol. Cell. Biol., 5:1188-1190.
- Gossen M. et al. (1992) Proc. Natl. Acad. Sci. USA. 89:5547-5551.
- Gossen M. et al. (1995) Science. 268:1766-1769.
- Graham et al. (1973) Virology 52:456-457.
- Green et al. (1986) Ann. Rev. Biochem. 55:569-597.
- 30 Greenspan and Bona, FASEB J. 7(5):437-444 (1989)
- Griffin et al. (1989) Science. 245:967-971.
- Grompe, M. (1993) Nature Genetics. 5:111-117.
- Grompe, M. et al. (1989) Proc. Natl. Acad. Sci. U.S.A. 86:5855-5892.
- Gu H. et al. (1993) Cell 73:1155-1164.
- 35 Guatelli J C et al. Proc. Natl. Acad. Sci. USA. 90:273-286.

- Hacia JG, Brody LC, Chee MS, Fodor SP, Collins FS, Nat Genet 1996;14(4):441-447
- Haff L. A. and Smirnov I. P. (1997) GenomeResearch, 7:378-388.
- Hames B.D. and Higgins S.J. (1985) Nucleic Acid Hybridization: A Practical Approach. Hames and Higgins Ed., IRL Press, Oxford.
- 5 Hammerling, et al., in: MONOCLONAL ANTIBODIES AND T-CELL HYBRIDOMAS 563-681 (Elsevier, N.Y., 1981)
- Harju L, Weber T, Alexandrova L, Lukin M, Ranki M, Jalanko A, Clin Chem 1993;39(11 Pt 1):2282-2287
- Harland et al. (1985) J. Cell. Biol. 101:1094-1095.
- 10 Harrop, J.A. et al. J. Immunol. 161(4): 1786-1794 (1998)
- Hawley M.E. et al. (1994) Am. J. Phys. Anthropol. 18:104.
- Henikoff and Henikoff, 1993, Proteins 17:49-61
- Higgins et al., 1996, Methods Enzymol. 266:383-402;
- Hillier L. and Green P. Methods Appl., 1991, 1: 124-8. Gu H. et al. (1994) Science 265:103-106.
- 15 Hoess et al. (1986) Nucleic Acids Res. 14:2287-2300.
- Huang L. et al. (1996) Cancer Res 56(5):1137-1141.
- Huston et al. Methods in Enzymology 203:46-88 (1991)
- Huygen et al. (1996) Nature Medicine. 2(8):893-898.
- 20 Izant J.G. and Weintraub H. (1984) Cell 36(4):1007-1015.
- Julan et al. (1992) J. Gen. Virol. 73:3251-3255.
- Kanegae Y. et al., Nucl. Acids Res. 23:3816-3821.
- Karlin and Altschul, 1990, Proc. Natl. Acad. Sci. USA 87:2267-2268;
- Kettleborough, C.A. et al. Eur. J. Immunol. 24:952-958 (1994)
- 25 Khoury J. et al. (1993) Fundamentals of Genetic Epidemiology, Oxford University Press, NY.
- Kim U-J. et al. (1996) Genomics 34:213-218.
- Klein et al. (1987) Nature. 327:70-73.
- Koller et al. (1992) Annu. Rev. Immunol. 10:705-730.
- Kozal MJ, Shah N, Shen N, Yang R, Fucini R, Merigan TC, Richman DD, Morris D, Hubbell E, Chee M, Gingeras TR, Nat Med 1996;2(7):753-759
- 30 Landegren U. et al. (1998) Genome Research, 8:769-776.
- Lander and Schork, Science, 265, 2037-2048, 1994
- Lenhard T. et al. (1996) Gene. 169:187-190.
- Liautaud, J. et al. Cytokine 9(4):233-241 (1997)
- 35 Linton M.F. et al. (1993) J. Clin. Invest. 92:3029-3037.

- Liu Z. et al. (1994) *Proc. Natl. Acad. Sci. USA.* 91: 4528-4262.
- Livak et al., *Nature Genetics*, 9:341-342, 1995
- Livak KJ, Hainer JW, *Hum Mutat* 1994;3(4):379-385
- Lockhart et al. (1996) *Nature Biotechnology* 14:1675-1680.
- 5 Mansour S.L. et al. (1988) *Nature.* 336:348-352.
- Marshall R. L. et al. (1994) *PCR Methods and Applications*. 4:80-84.
- McCormick et al. (1994) *Genet. Anal. Tech. Appl.* 11:158-164.
- McLaughlin B.A. et al. (1996) *Am. J. Hum. Genet.* 59:561-569.
- Morton N.E. (1955) *Am. J. Hum. Genet.* 7:277-318.
- 10 Morrison, *Science* 229:1202 (1985)
- Muller, Y.A. et al. *Structure* 6(9):1153-1167(1998)
- Mullinax, R.L. et al. *BioTechniques* 12(6):864-869(1992)
- Muzyczka et al. (1992) *Curr. Topics in Micro. and Immunol.* 158:97-129.
- Padlan E.A., *Molecular Immunology* 28(4/5): 489-498(1991)
- 15 Nada S. et al. (1993) *Cell* 73:1125-1135.
- Nagy A. et al. (1993) *Proc. Natl. Acad. Sci. USA.* 90: 8424-8428.
- Naramura, M. et al. *Immunol. Lett.* 39:91-99(1994)
- Narang SA, Hsiung HM, Brousseau R, *Methods Enzymol* 1979;68:90-98
- Neda et al. (1991) *J. Biol. Chem.* 266:14143-14146.
- 20 Newton et al. (1989) *Nucleic Acids Res.* 17:2503-2516.
- Nickerson D.A. et al. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87:8923-8927.
- Nicolau et al. (1982) *Biochim. Biophys. Acta.* 721:185-190.
- Nissinoff, J. *Immunol.* 147(8): 2429-2438 (1991).
- Nyren P, Pettersson B, Uhlen M, *Anal Biochem* 1993;208(1):171-175
- 25 O'Reilly et al. (1992) *Baculovirus Expression Vectors: A Laboratory Manual.* W. H. Freeman and Co., New York.
- Ohno et al. (1994) *Science.* 265:781-784.
- Oi et al., *BioTechniques* 4:214 (1986)
- Orita et al. (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86: 2776-2770.
- 30 Ott J. (1991) *Analysis of Human Genetic Linkage.* John Hopkins University Press, Baltimore.
- Pastinen et al., *Genome Research* 1997; 7:606-614
- Pearson and Lipman, 1988, *Proc. Natl. Acad. Sci. USA* 85(8):2444-2448;
- Pease S. and William R.S. (1990) *Exp. Cell. Res.* 190:09-211.
- Perlin et al. (1994) *Am. J. Hum. Genet.* 55:777-787.
- 35 Persic, L. et al. *Gene* 187 9-18(1997)

- Peterson et al.(1993) Proc. Natl. Acad. Sci. USA. 90: 7593-7597.
- Pietu et al.(1996) Genome Research.6:492-503.
- Pitard, V. et al. J. Immunol. Methods 205(2):177-190(1997)
- Potter et al. (1984) Proc. Natl.Acad. Sci. U.S.A. 81(22):7161-7165.
- 5 Prat, M. et al. J. Cell. Sci. 111(Pt2): 237-247(1998)
- Rayl et al., (1996) J. Bio.Chem, 271, 2225-2233.
- Reid L.H. et al. (1990) Proc. Natl. Acad.Sci. U.S.A. 87:4299-4303.
- Risch, N. and Merikangas, K. (1996)Science. 273:1516-1517.
- Robertson E. (1987) "Embryo-Derived StemCell Lines." In: E.J. Robertson Ed.
- 10 Teratocarcinomas And Embrionic Stem Cells: A Practical Approach. IRL Press, Oxford, pp. 71.
- Roguska M.A. et al. PNAS 91:969-973), (1994)
- Rossiet al. (1991) Pharmacol. Ther. 50:245-254.
- Roth J.A. et al. (1996)Nature Medicine. 2(9):985-991.
- 15 Roux et al. (1989) Proc. Natl. Acad.Sci. U.S.A. 86:9079-9083.
- Ruano et al. (1990) Proc. Natl. Acad.Sci. U.S.A. 87:6296-6300.
- Sambrook, J., Fritsch, E.F., and T.Maniatis. (1989) Molecular Cloning: A Laboratory Manual. 2ed. ColdSpring Harbor Laboratory, Cold Spring Harbor, New York.
- Samson M, etal. (1996) Nature, 382(6593):722-725.
- 20 Samulski et al. (1989) J.Virol. 63:3822-3828.
- Sanchez-Pescador R. (1988) J. Clin.Microbiol. 26(10):1934-1938.
- Sarkar, G. and Sommer S.S. (1991)Biotechniques.
- Sauer B. et al. (1988) Proc. Natl. Acad. Sci.U.S.A. 85:5166-5170.
- Sawai, H. et al. AJRI 34:26-34(1995)
- 25 Schaid D.J. et al. (1996) Genet. Epidemiol.13:423-450.
- Schedl A. et al. (1993a) Nature. 362:258-261.
- Schedl et al. (1993b) Nucleic Acids Res. 21:4783-4787.
- Schedl etal. (1993b) Nucleic Acids Res. 21:4783-4787.
- Schena et al. (1995)Science. 270:467-470.
- 30 Schena et al. (1996) Proc. Natl. Acad. Sci.U.S.A. 93(20):10614-10619.
- Schneider et al.(1997) Arlequin: ASoftware For Population Genetics Data Analysis. University of Geneva.
- Schwartz and Dayhoff, eds., 1978, Matrices for Detecting DistanceRelationships: Atlas of Protein Sequence and Structure, Washington:National Biomedical Research Foundation
- 35 Sczakiel G. et al. (1995)Trends Microbiol. 3(6):213-217.

- Shay J.W. et al. (1991) *Biochem. Biophys. Acta.* 1072:1-7.
- Sheffield, V.C. et al. (1991) *Proc. Natl. Acad. Sci. U.S.A.* 49:699-706.
- Shizuya et al. (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89:8794-8797.
- Shoemaker DD, Lashkari DA, Morris D, Mittmann M, Davis RW, *Nat Genet* 1996;14(4):450-456
- 5 Shu, L. et al. *PNAS* 90:7995-7999(1993)
- Skerra, A. et al. *Science* 240:1038-1040. (1988)
- Smith(1957) *Ann. Hum. Genet.* 21:254-276.
- Smith et al. (1983) *Mol. Cell. Biol.* 3:2156-2165.
- 10 Sosnowski R.G. et al. (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94:1119-1123.
- Spielmann S. and Ewens W. J. (1998) *Am. J. Hum. Genet.* 62:450-458.
- Spielmann S. et al. (1993) *Am. J. Hum. Genet.* 52:506-516.
- Sternberg N.L. (1992) *Trends Genet.* 8:1-16.
- Sternberg N.L. (1994) *Mamm. Genome.* 5:397-404.
- 15 Studnicka G.M. et al. *Protein Engineering* 7(6):805-814(1994)
- Syvanen AC, *Clin Chim Acta* 1994;226(2):225-236
- Tacson et al. (1996) *Nature Medicine.* 2(8):888-892.
- Taryman, R.E. et al. *Neuron* 14(4):755-762(1995)
- Te Riele et al. (1990) *Nature.* 348:649-651.
- 20 Terwilliger J.D. and Ott J. (1994) *Handbook of Human Genetic Linkage.* John Hopkins University Press, London.
- Thomas K.R. et al. (1986) *Cell.* 44:419-428.
- Thomas K.R. et al. (1987) *Cell.* 51:503-512.
- Thompson et al., 1994, *Nucleic Acids Res.* 22(2):4673-4680;
- 25 Tur-Kaspa et al. (1986) *Mol. Cell. Biol.* 6:716-718.
- Tyagi et al. (1998) *Nature Biotechnology.* 16:49-53.
- Urdea M.S. (1988) *Nucleic Acids Research.* 11:4937-4957.
- Urdea M.S. et al. (1991) *Nucleic Acids Symp. Ser.* 24:197-200.
- Van der Lugt et al. (1991) *Gene.* 105:263-267.
- 30 Vil, H. et al. *PNAS* 89:11337-11341(1992)
- Vlasak R. et al. (1983) *Eur. J. Biochem.* 135:123-126.
- Wabiko et al. (1986) *DNA.* 5(4):305-314.
- Walker et al. (1996) *Clin. Chem.* 42:9-13.
- Weir, B.S. (1996) *Genetic data Analysis II: Methods for Discrete population genetic Data,*
- 35 Sinauer Assoc., Inc., Sunderland, MA, U.S.A.

- White, M.B. et al. (1992) *Genomics*. 12:301-306.
- White, M.B. et al. (1997) *Genomics*. 12:301-306.
- Wong et al. (1980) *Gene*. 10:87-94.
- Wood S.A. et al. (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:4582-4585.
- 5 Wu and Wu (1987) *J. Biol. Chem.* 262:4429-4432.
- Wu and Wu (1988) *Biochemistry*. 27:887-892.
- Wu et al. (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86:2757.
- Yagi T. et al. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87:9918-9922.
- Yoon, D.Y. et al. *J. Immunol.* 160(7): 3170-3179(1998)
- 10 Zhao et al. (1998) *Am.J. Hum. Genet.* 63:225-240.
- Zheng, X.X. et al. *J. Immunol.* 154:5590-5600(1995)
- Zhu, Z. et al. *Cancer Res.* 58(15): 3209-3214(1998)
- Zou Y. R. et al. (1994) *Curr. Biol.* 4:1099-1103

## CLAIMS

1. An isolated, purified or recombinant polynucleotide comprising a contiguous span of at least 30 nucleotides selected from the group consisting of:
- 5 (a) SEQ ID Nos. 2 to 17, and the complements thereof;
- (b) SEQ ID Nos 24, 25, 28, 74 and 79 to 132, and the complements thereof; and
- (c) at least one of nucleotide positions 199122 to 249803 of SEQ ID No 1.
2. An isolated, purified or recombinant polynucleotide according to claim 1, wherein said
- 10 contiguous span of SEQ ID No 1 or the complements thereof comprises at least one of the following nucleotide positions of SEQ ID No 1: 201123 to 201234, 201123 to 201560, 214676 to 214793, 215702 to 215746, 216836 to 216994, 216836 to 217077, 217671 to 217764, 227655 to 227736, 238715 to 238919, 240440 to 240673, 240440 to 241153, 241072 to 241291, 244353 to 244561 and 246273 to 247802;
- 15 3. An isolated, purified or recombinant polynucleotide according to claim 1, wherein said span comprises a biallelic marker selected from the group consisting of A13 to A18, A20 to A46, A49 to A52, A55, A57, A59 to A63, A72 to A73, A76 and A123 .
- 20 4. A recombinant vector comprising a polynucleotide according to any one of claims 1 to 3.
5. A host cell comprising a recombinant vector according to claim 4.
6. A non-human host animal or mammal comprising a recombinant vector according to claim
- 25 5.
7. A mammalian host cell comprising a g35030 gene disrupted by homologous recombination with a knock out vector, comprising a polynucleotide according to any one of claims 1 to 3.
- 30 8. A non-human host mammal comprising a g35030 gene disrupted by homologous recombination with a knock out vector, comprising a polynucleotide according to any one of claims 1 to 3.
9. Use of a polynucleotide comprising a contiguous span of at least 12 nucleotides of a sequence
- 35 selected from the group consisting of SEQ ID Nos. 2 to 17, SEQ ID Nos 24, 25, 28, 74 and 79

to 132, at least one of nucleotide positions 199122 to 249803 of SEQ ID No 1, and the complementary sequences thereto, for determining the identity of the nucleotide at a biallelic marker selected from the group consisting of A13 to A18, A20 to A46, A49 to A52, A55, A57, A59 to A63, A72 to A73, A76 and A123 .

5

10. Use of a polynucleotide according to claim 9 in a microsequencing assay, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide and wherein the 3' end of said polynucleotide is located 1 nucleotide upstream of a biallelic marker in said sequence.

10

11. Use of a polynucleotide according to claim 10 in a hybridization assay, wherein said span includes a biallelic marker.

12. A polynucleotide according to any one of claims 1 to 3 attached to a solid support.

15

13. An array of polynucleotides comprising at least one polynucleotide according to claim 12.

14. An array according to claim 13, wherein said array is addressable.

20

15. A polynucleotide according to any one of claims 1 to 3 or 12 further comprising a label.

16. A method of genotyping comprising determining the identity of a nucleotide at a biallelic marker selected from the group consisting of A13 to A18, A20 to A46, A49 to A52, A55, A57, A59 to A63, A72 to A73, A76 and A123 or the complement thereof in a biological sample.

25

17. A method according to claim 16, wherein said biological sample is derived from a single subject.

18. A method according to claim 16, wherein said biological sample is derived from multiple subjects.

30

19. A method according to claim 16, further comprising amplifying a portion of said sequence comprising the biallelic marker prior to said determining step.

35

20. A method according to claim 19, wherein said amplifying is performed by PCR.



21. A method according to claim 16, wherein said determining is performed by a hybridization assay.
- 5 22. A method according to claim 16, wherein said determining is performed by a sequencing assay.
23. A method according to claim 16, wherein said determining is performed by a microsequencing assay.
- 10 24. A method according to claim 16, wherein said determining is performed by an enzyme-based mismatch detection assay.
25. A method of estimating the frequency of an allele of a biallelic marker in a population comprising:
- 15 a) genotyping individuals from said population for said biallelic marker according to the method of claim 16; and
- b) determining the proportional representation of said biallelic marker in said population.
- 20 26. A method of detecting an association between a genotype and a trait, comprising the steps of:
- a) determining the frequency of at least one biallelic marker in trait positive population according to the method of claim 25;
- b) determining the frequency of at least one biallelic marker in a control population according to
- 25 the method of claim 25; and
- c) determining whether a statistically significant association exists between said genotype and said trait.
27. A method of estimating the frequency of a haplotype for a set of biallelic markers in a
- 30 population, comprising:
- a) genotyping at least one biallelic marker according to claim 16 for each individual in said population;
- b) genotyping a second biallelic marker by determining the identity of the nucleotides at said second biallelic marker for both copies of said second biallelic marker present in the
- 35 genome of each individual in said population; and

c) applying a haplotype determination method to the identities of the nucleotides determined in steps a) and b) to obtain an estimate of said frequency.

5 28. A method according to claim 27, wherein said haplotype determination method is selected from the group consisting of asymmetric PCR amplification, double PCR amplification of specific alleles, the Clark method, or an expectation maximization algorithm.

29. A method of detecting an association between a haplotype and a trait, comprising the steps of:

10 a) estimating the frequency of at least one haplotype in a trait positive population according to the method of claim 27;

b) estimating the frequency of said haplotype in a control population according to the method of claim 27; and

15 c) determining whether a statistically significant association exists between said haplotype and said trait.

30. A method according to claim 26, wherein said genotyping of step a) is performed on each individual of said population.

20 31. A method according to claim 26, wherein said genotyping is performed on a single pooled biological sample derived from said population.

32. A method of detecting an association between an allele and a phenotype, comprising the steps of:

25 a) determining the frequency of at least one biallelic marker allele in a trait positive population according to the method of claim 25;

b) determining the frequency of said biallelic marker allele in a control population according to the method of claim 25; and

30 c) determining whether a statistically significant association exists between said allele and said phenotype.

33. A method according to claim 26 or 29, wherein said trait is schizophrenia or bipolar disorder.

35 34. A method according to claim 26 or 29, wherein said trait is predisposition to schizophrenia

or bipolar disorder, an early onset of schizophrenia or bipolar disorder, or a beneficial response to or side effects related to treatment against schizophrenia or bipolar disorder.

5 35. A method according to claim 39, wherein said phenotype is a symptom of schizophrenia or bipolar disorder.

36. A method according to claim 26 or 29, wherein said control population is a trait negative population.

10 37. A method according to claim 26 or 29, wherein said case control population is a random population.

38. A method of determining whether an individual is at risk of schizophrenia or bipolar disorder, comprising:

- 15 a) genotyping at least one biallelic marker according to the method of claim 16; and  
b) correlating the result of step a) with a risk of developing schizophrenia or bipolar disorder.

39. A diagnostic kit comprising a polynucleotide according to any one of claims 1 to 3 or 12.

20 40. A purified or isolated g35030 polypeptide which is encoded by a nucleic acid comprising a nucleotide sequence selected from the group consisting of SEQ ID Nos 2 to 17, and fragments or variants thereof.

25 41. A purified or isolated g35030 polypeptide comprising at least 6 contiguous amino acid residues of any of SEQ ID Nos 18 to 23.

42. A purified or isolated g35030 polypeptide according to claim 41, comprising at least one amino acid substitution, addition or deletion.

30 43. A method for producing a g35030 polypeptide, wherein said method comprises the following steps:  
a) providing a cell host comprising a recombinant vector according to claim 4 containing a nucleic acid encoding a g35030 polypeptide;  
b) recovering the g35030 polypeptide produced by said recombinant cell host.

35

44. The method according to claim 43, wherein the recombinant cell host is a recombinant cell host according to claim 5.
- 5 45. An isolated or purified antibody composition capable of selectively binding to a polypeptide according to claim 40 to 42.
46. A method for specifically detecting the presence of a g35030 polypeptide in a biological sample, said method comprising the following steps :
- 10 a) bringing into contact the biological sample with an antibody directed against a g35030 polypeptide according to any one of claims 40 to 42;
- b) detecting the antigen-antibody complex formed between said antibody and said polypeptide.
47. A diagnostic kit for detecting *in vitro* the presence of a g35030 polypeptide in a biological sample, said kit comprising:
- 15 a) a polyclonal or monoclonal antibody directed against a g35030 polypeptide according to any one of claims 40 to 42 or a fragment thereof, optionally labeled;
- b) a reagent allowing the detection of the antigen-antibody complexes formed between said g35030 polypeptide and an antibody.
- 20 48. A method for the screening of a candidate substance, wherein said method comprises the following steps:
- a) providing a polypeptide according to anyone of claims 40 to 42;
- b) obtaining a candidate substance;
- c) bringing into contact said polypeptide with said candidate substance;
- 25 d) detecting the complexes formed between said polypeptide and said candidate substance.
49. The method of claim 48, wherein at step d), the complexes formed are incubated in the presence of a polyclonal or a monoclonal antibody according to claim 45 .
- 30 50. A kit for screening a candidate substance interacting with a g35030 polypeptide, wherein said kit comprises:
- a) a polypeptide according to anyone of claims 40 to 42;
- b) optionally a monoclonal or a polyclonal antibody according to claim 45.
- 35 51. A method for the screening of a candidate substance, where said method comprises the

following steps:

- a) cultivating a prokaryotic or an eukaryotic cell that has been transfected with a nucleotide sequence encoding a g35030 protein or a variant or a fragment thereof, placed under the control of its own promoter;
- 5 b) bringing into contact the cultivated cell with a molecule to be tested;
- c) quantifying the expression of the g35030 protein or a variant or a fragment thereof.

52 A computer readable medium having stored thereon a sequence selected from the group consisting of a nucleic acid code comprising one of the following:

- 10 a) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, 1000 or 2000 nucleotides of SEQ ID No. 1, and the complements thereof, wherein said contiguous span comprises at least one of the following nucleotide positions of SEQ ID No 1: 201123 to 247802 and 199122 to 249803;
- b) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 15 200, 500, 1000 or 2000 nucleotides of any of SEQ ID Nos. 2 to 17, and the complements thereof, to the extent that such a length is consistent with the particular sequence ID;
- c) a contiguous span of at least 8, 12, 15, 18, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100 or 200 nucleotides, to the extent that such a length is consistent with the particular sequence ID, of SEQ ID Nos. 24, 25, 28, 74 and 79 to 132, or the complements thereof;
- 20 d) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90 or 100 nucleotides of SEQ ID No. 1 or the complements thereof wherein said contiguous span comprises at least one of the following nucleotide positions of SEQ ID No 1: 201123 to 201234, 201123 to 201560, 214676 to 214793, 215702 to 215746, 216836 to 216994, 216836 to 217077, 217671 to 217764, 227655 to 227736, 238715 to 238919, 240440 to 240673, 25 240440 to 241153, 241072 to 241291, 244353 to 244561 and 246273 to 247802;
- e) a contiguous span according to a), b), c) or d), wherein said span includes a biallelic marker selected from the group consisting of A1 to A127;
- f) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, 1000 or 2000 nucleotides of any of SEQ ID No 1, and the complements thereof, 30 wherein said span includes a polymorphism selected from the group consisting of A13 to A18, A20 to A46, A49 to A52, A55, A57, A59 to A63, A72 to A73, A76 and A123 ; and
- g) a nucleotide sequence complementary to any one of the contiguous spans of a), b), c), d), e), f), and g).

- 35 53. A computer readable medium having stored thereon a sequence consisting of a polypeptide

code comprising a contiguous span of at least 6 amino acids of a polypeptide sequence selected from the group consisting of SEQ ID Nos. 18 to 23.

5 54. A computer system comprising a processor and a data storage device wherein said data storage device comprises a computer readable medium according to any one of claims 52 or 53.

55. A computer system according to claim 54, further comprising a sequence comparer and a data storage device having reference sequences stored thereon.

10 56. A computer system of claim 55 wherein said sequence comparer comprises a computer program which indicates polymorphisms.

57. A computer system of claim 55 further comprising an identifier which identifies features in said sequence.

15

58. A method for comparing a first sequence to a reference sequence, comprising the steps of:

a) reading said first sequence and said reference sequence through use of a computer program which compares sequences; and

20 b) determining differences between said first sequence and said reference sequence with said computer program,

wherein said first sequence is selected from the group consisting of a nucleic acid code comprising one of the following:

25 (i) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, 1000 or 2000 nucleotides of SEQ ID No. 1, and the complements thereof, wherein said contiguous span comprises at least one of the following nucleotide positions of SEQ ID No 1: 201123 to 247802 and 199122 to 249803;

(ii) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, 1000 or 2000 nucleotides of any of SEQ ID Nos. 2 to 17, and the complements thereof, to the extent that such a length is consistent with the particular sequence ID;

30 (iii) a contiguous span of at least 8, 12, 15, 18, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100 or 200 nucleotides, to the extent that such a length is consistent with the particular sequence ID, of SEQ ID Nos. 24, 25, 28, 74 and 79 to 132, or the complements thereof;

35 (iv) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90 or 100 nucleotides of SEQ ID No. 1 or the complements thereof wherein said contiguous span comprises at least one of the following nucleotide positions of SEQ ID No 1: 201123 to

201234, 201123 to 201560, 214676 to 214793, 215702 to 215746, 216836 to 216994, 216836 to 217077, 217671 to 217764, 227655 to 227736, 238715 to 238919, 240440 to 240673, 240440 to 241153, 241072 to 241291, 244353 to 244561 and 246273 to 247802;

5 (v) a contiguous span according to (i) to (iv), wherein said span includes a biallelic marker selected from the group consisting of A1 to A127.

(vi) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, 1000 or 2000 nucleotides of any of SEQ ID No 1, and the complements thereof, wherein said span includes a polymorphism selected from the group consisting of A13 to A18, A20 to A46, A49 to A52, A55, A57, A59 to A63, A72 to A73, A76 and A123;

10 (vii) a nucleotide sequence complementary to any one of the contiguous spans of (i) to (vi); and

(viii) a polypeptide code comprising a contiguous span of at least 6 amino acids of a polypeptide sequence selected from the group consisting of SEQ ID Nos 18 to 23.

15 59. A method according to claim 58, wherein said step of determining differences between the first sequence and the reference sequence comprises identifying at least one polymorphism.

60. A method for identifying a feature in a sequence, comprising the steps of:

20 a) reading said sequence through the use of a computer program which identifies features in sequences; and

b) identifying features in said sequence with said computer program;

wherein said sequence is selected from the group consisting of a nucleic acid code comprising one of the following:

25 (i) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, 1000 or 2000 nucleotides of SEQ ID No. 1, and the complements thereof, wherein said contiguous span comprises at least one of the following nucleotide positions of SEQ ID No 1: 201123 to 247802 and 199122 to 249803;

30 (ii) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, 1000 or 2000 nucleotides of any of SEQ ID Nos. 2 to 17, and the complements thereof, to the extent that such a length is consistent with the particular sequence ID;

(iii) a contiguous span of at least 8, 12, 15, 18, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100 or 200 nucleotides, to the extent that such a length is consistent with the particular sequence ID, of SEQ ID Nos. 24, 25, 28, 74 and 79 to 132, or the complements thereof;

35 (iv) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90 or 100 nucleotides of SEQ ID No. 1 or the complements thereof wherein said contiguous

span comprises at least one of the following nucleotide positions of SEQ ID No 1: 201123 to 201234, 201123 to 201560, 214676 to 214793, 215702 to 215746, 216836 to 216994, 216836 to 217077, 217671 to 217764, 227655 to 227736, 238715 to 238919, 240440 to 240673, 240440 to 241153, 241072 to 241291, 244353 to 244561 and 246273 to 247802;

5 (v) a contiguous span according to (i) to (iv), wherein said span includes a biallelic marker selected from the group consisting of A1 to A127;

(vi) a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, 1000 or 2000 nucleotides of any of SEQ ID No 1, and the complements thereof, wherein said span includes a polymorphism selected from the group consisting of A13 to A18, A20 to A46, A49 to A52, A55, A57, A59 to A63, A72 to A73, A76 and A123 ;

10 (vii) a nucleotide sequence complementary to any one of the contiguous spans of (i) to (vi); and

(viii) a polypeptide code comprising a contiguous span of at least 6 amino acids of a polypeptide sequence selected from the group consisting of SEQ ID Nos 18 to 23.

15



1/15

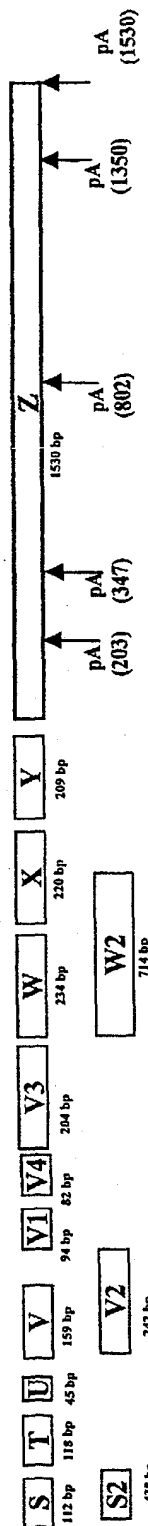


Figure 1

2/15

# SIGNIFICANCE OF SINGLE BIALLELIC MARKER ASSOCIATION STUDY IN THE 13q31-33 REGION

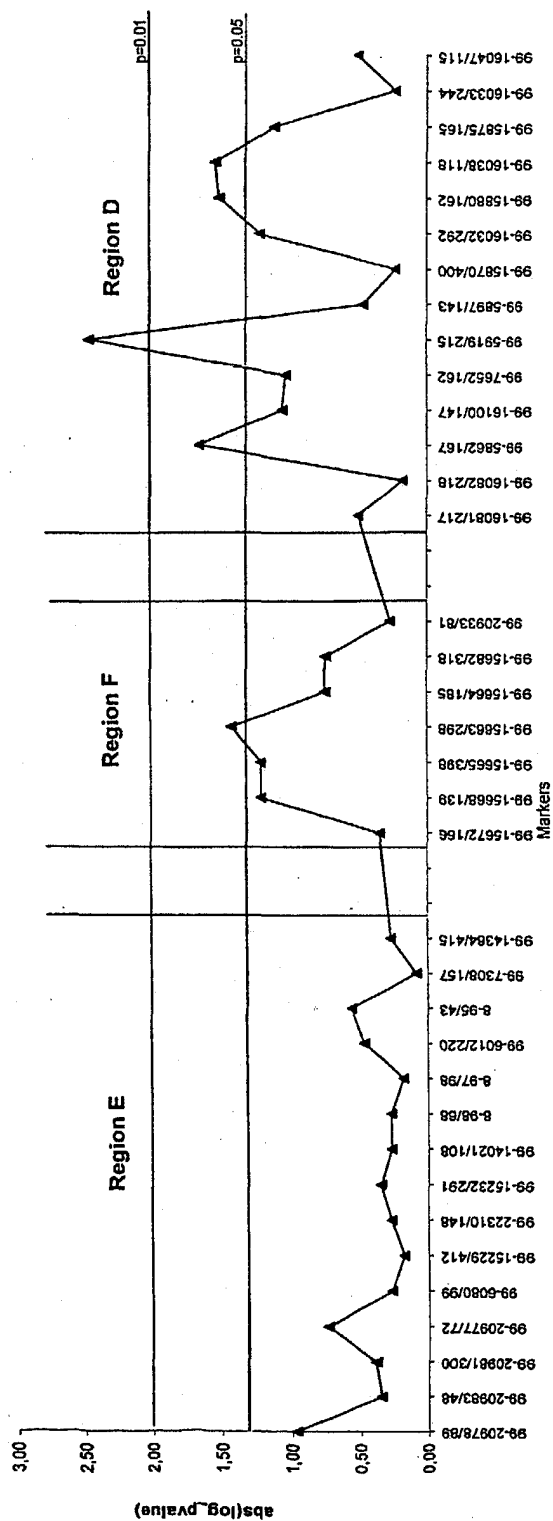


Figure 2

Detailed haplotype analysis (case/controls: N=215/N=241)

[illegible]

**maxM:** Table of haplotypes giving the Maximum positive difference between cases/controls.

- 1: Frequency of the haplotype leading to the maximum chi square test
- 2: Test on the frequency of this haplotype in cases vs in controls
- 3: *n*-value assuming that the test has a chi-square distribution with 1 degree of freedom

### Figure 3

4/15

Figure 4

Markers involved in selected haplotypes

			markers											
# of markers in the haplotype	Threshold of significance	# of haplotypes concerned	99-16082/218	99-5862/167	99-16100/147	99-7652/162	99-5919/215	99-5897/143	99-16032/292	99-15880/162	99-16038/118	99-15875/165	99-16033/244	99-16047/115
2-mks	$\leq 10^{-2}$	7	*	*	*	*	*	*	*	*	*	*		
3-mks	$\leq 10^{-5}$	10	*	*	*	*	*	*	*	*	*	*	*	
4-mks	$\leq 10^{-6}$	5	*	*	*	*	*	*	*		*		*	*

5/15

## SCHIZOPHRENIA STUDY

MARKERS	HAPLOTYPE DESCRIPTION										HAPLOTYPE TEST BY PERMUTATION PROCEDURE			
	Statistics on a haplotype $\alpha$										av.Chiz	# Hets/ops		
	Frequency $f_i$													
	Cases OR Controls													
sample size (cases/controls)	134/136	137/137	139/135	140/140	141/141	142/142	143/143	144/144	145/145	146/146	147/147	148/148		
frequency in controls (Allele)	45.95 (G)	59.48 (T)	50 (T)	49.64 (G)	51.11 (G)	50.71 (T)	60.28 (A)	61.07 (C)	59.49 (G)	38.93 (C)	62.95 (C)	56.38 (G)		
diff (cases-controls)	5.2	1.5	4.7	4.6	4.2	5.7	9.2	3.2	3.3	3.6	4.9	9.2		
p-value	0.221	0.655	0.254	0.254	0.317	0.168	0.0213	0.403	0.752	0.371	0.221	0.0238		
HW-test	0 (NS)	-0.007 (NS)	-0.009 (NS)	-0.023 (NS)	-0.006 (NS)	-0.013 (NS)	-0.029 (NS)	-0.013 (NS)	0.012 (NS)	-0.023 (NS)	0.011 (NS)	-0.019 (NS)		
controls	140 vs 136	140 vs 138	139 vs 138	137 vs 133	132 vs 133	135 vs 132	140 vs 137	140 vs 138	140 vs 139	140 vs 140	140 vs 141	140 vs 142		
haplotype 1	T	T	T	T	T	T	T	T	T	T	T	T		
haplotype 2														
haplotype 3														
haplotype 4														
haplotype 5														
haplotype 6														
haplotype 8														
haplotype 9														
haplotype 154														
haplotype 155														
haplotype 156														
haplotype 157														
haplotype 158														
haplotype 159														
haplotype 160														
haplotype 161														
haplotype 162														
haplotype 163														
haplotype 970														
haplotype 971														
haplotype 972														
haplotype 973														

Figure 5

MARKERS	HAPLOTYPE DESCRIPTION													HAPLOTYPE TEST BY PERMUTATION PROCEDURE																											
	Statistics on a haplotype <sup>(2)</sup>													Max Ch2	# het/cbs																										
																Frequency <sup>(1)</sup>																									
																													Controls												
OR													p-value (adj) <sup>(3)</sup>																												
Controls													Cases		OR		p-value (adj) <sup>(3)</sup>																								
Cases													OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																						
HWT-test													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
controls													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
cases													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
sample size (cases/controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
frequency in controls (Allele)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
HWT-test													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
controls													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
cases													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
sample size (cases/controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
frequency in controls (Allele)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
HWT-test													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
controls													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
cases													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
sample size (cases/controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
frequency in controls (Allele)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
HWT-test													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
controls													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
cases													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
sample size (cases/controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
frequency in controls (Allele)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
HWT-test													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
controls													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
cases													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
sample size (cases/controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
frequency in controls (Allele)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
HWT-test													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
controls													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
cases													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
sample size (cases/controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
frequency in controls (Allele)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
HWT-test													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
controls													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
cases													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
sample size (cases/controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
frequency in controls (Allele)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
HWT-test													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
controls													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
cases													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
sample size (cases/controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
frequency in controls (Allele)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
HWT-test													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
controls													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
cases													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
sample size (cases/controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
frequency in controls (Allele)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
HWT-test													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
controls													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
cases													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
sample size (cases/controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
frequency in controls (Allele)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
HWT-test													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
controls													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
cases													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
sample size (cases/controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
frequency in controls (Allele)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
HWT-test													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
controls													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
cases													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
sample size (cases/controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
frequency in controls (Allele)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
HWT-test													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
controls													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
cases													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
sample size (cases/controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
frequency in controls (Allele)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
HWT-test													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
controls													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
cases													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
sample size (cases/controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
frequency in controls (Allele)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
HWT-test													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
controls													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
cases													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
sample size (cases/controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
frequency in controls (Allele)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
HWT-test													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
controls													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
cases													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
sample size (cases/controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
frequency in controls (Allele)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
HWT-test													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
controls													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
cases													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
sample size (cases/controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
frequency in controls (Allele)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
HWT-test													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
controls													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
cases													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
sample size (cases/controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
frequency in controls (Allele)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
diff (cases-controls)													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
p-value													Cases		OR		p-value (adj) <sup>(3)</sup>		Max Ch2		# het/cbs																				
HWT-test													Cases		OR		p-value (adj) <sup>(3)</sup>																								

Figure 6

# SCHIZOPHRENIA STUDY OMNIBUS LR TEST VALUES DISTRIBUTIONS ALL SAMPLE

CASES (N=213) / ALL CONTROLS (N=241)

Combination of 2 markers

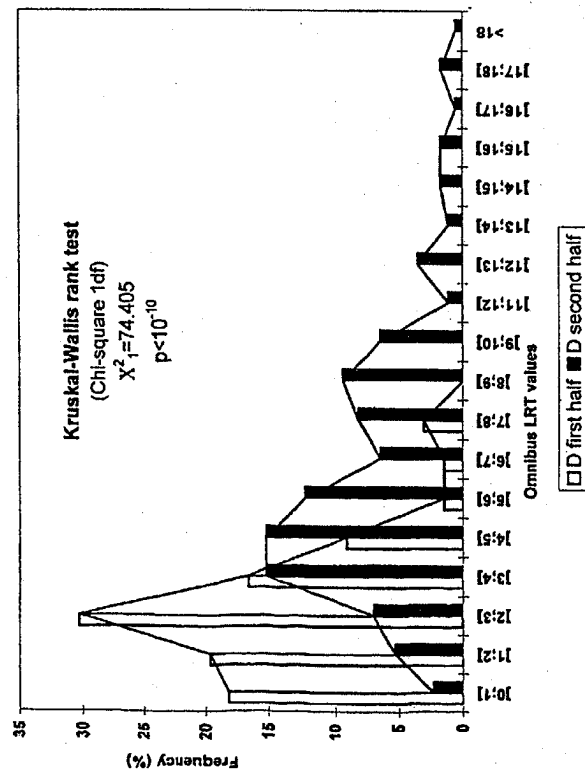


Figure 7A

Combination of 3 markers

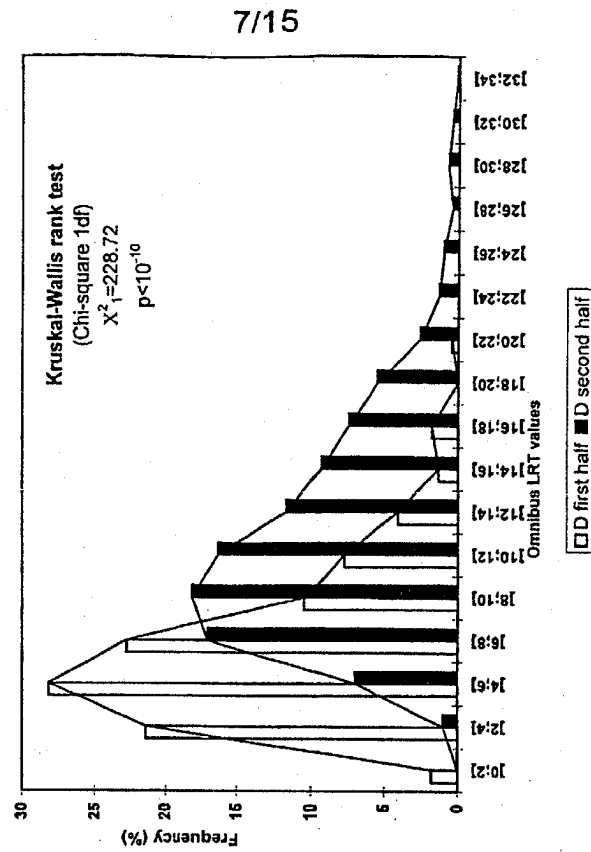


Figure 7B

# SCHIZOPHRENIA STUDY HAPLO\_MAXM TEST DISTRIBUTIONS ALL SAMPLE

CASES (N=213) / ALL CONTROLS (N=241)

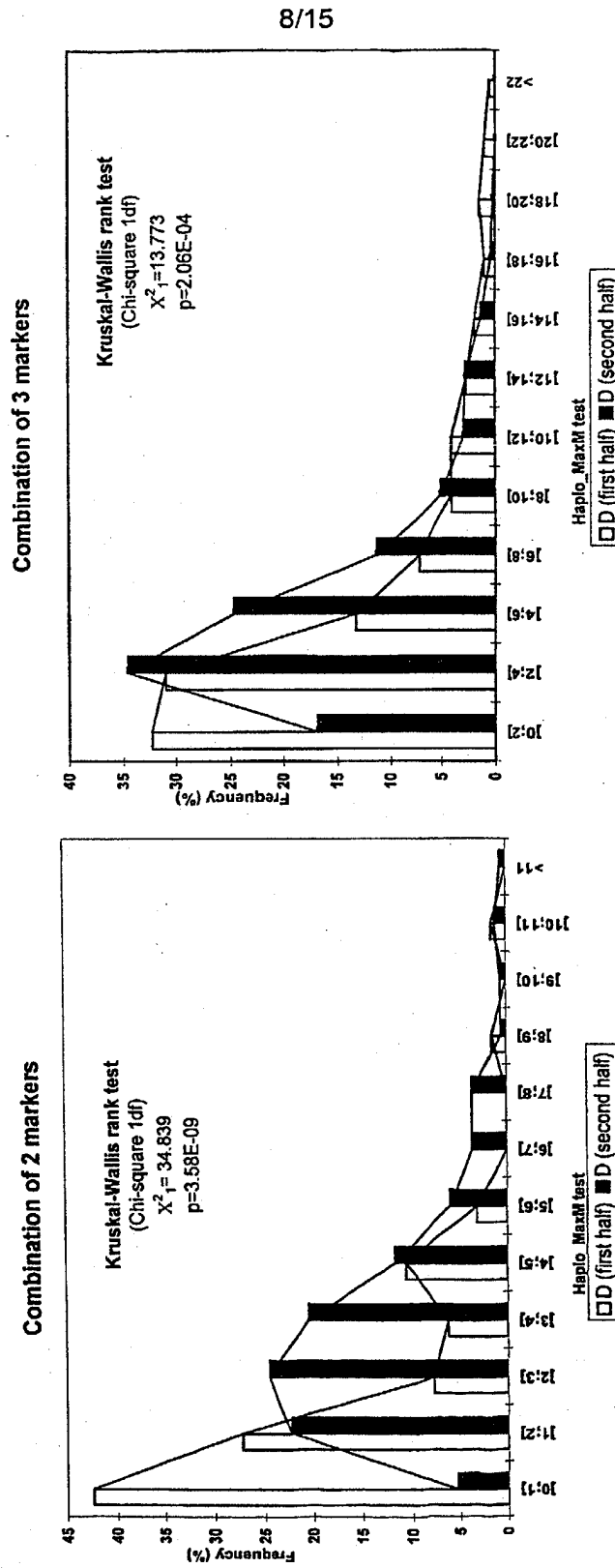


Figure 8A

Figure 8B



# BIPOLAR DISORDERS STUDY OMNIBUS LR TEST VALUES DISTRIBUTIONS FIRST HALF AND SECOND HALF D REGION

CAUCASIAN CASES (N=182) / CAUCASIAN CONTROLS (N=177)

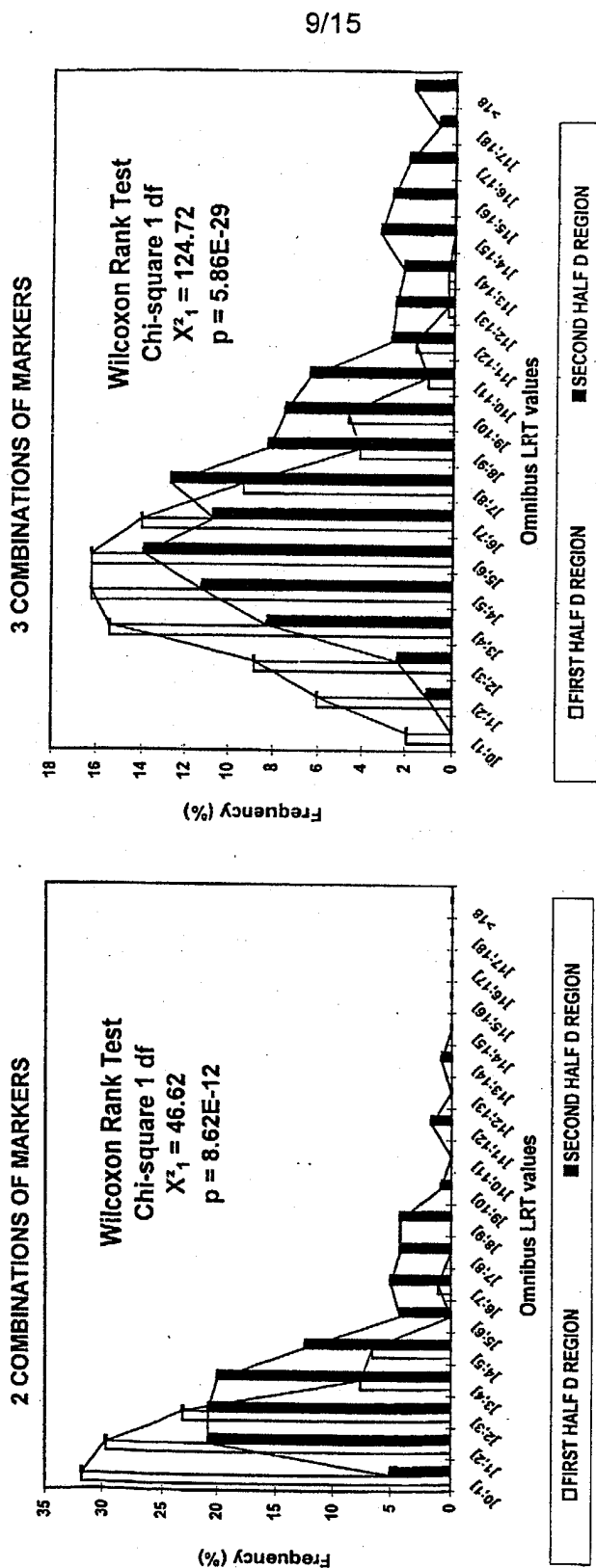


Figure 9A

Figure 9B

# BIPOLAR DISORDERS STUDY HAPLO\_MAXM TEST VALUES DISTRIBUTIONS

CAUCASIAN CASES (N=182) / CAUCASIAN CONTROLS (N=177)

Combination of 2 markers

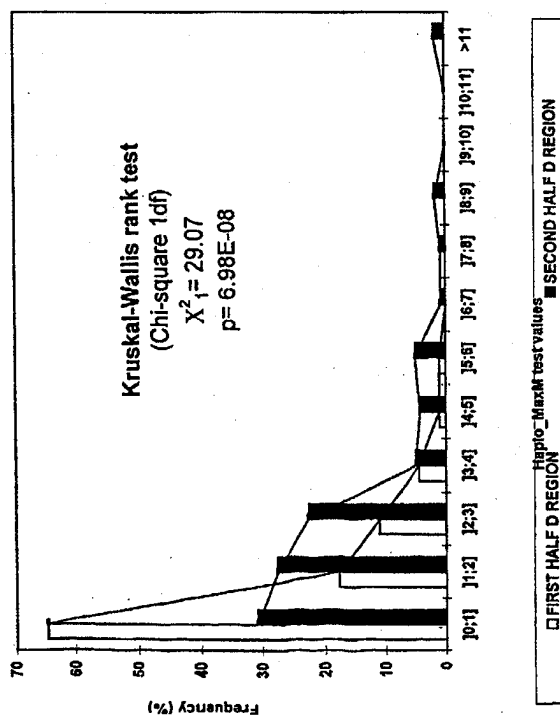


Figure 10A

10/15

Combination of 3 markers

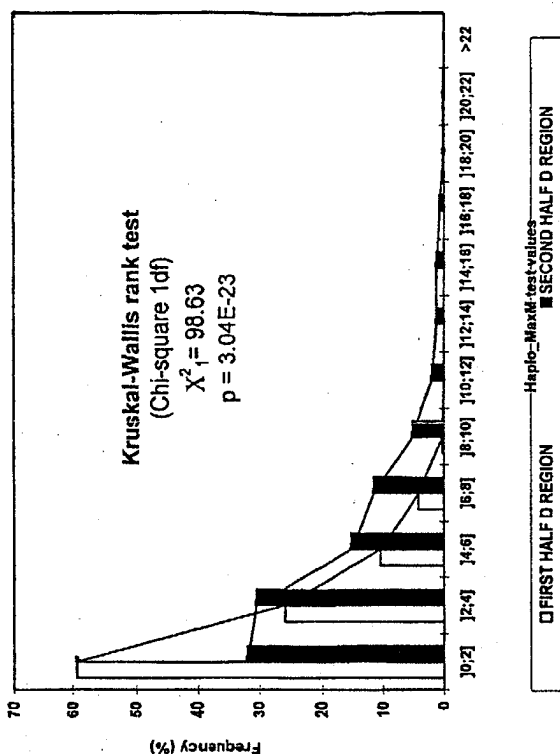


Figure 10B

# BIPOLAR DISORDERS STUDY HAPLO\_MAXS TEST VALUES DISTRIBUTIONS

CAUCASIAN CASES (N=182) / CAUCASIAN CONTROLS (N=177)

Combination of 2 markers

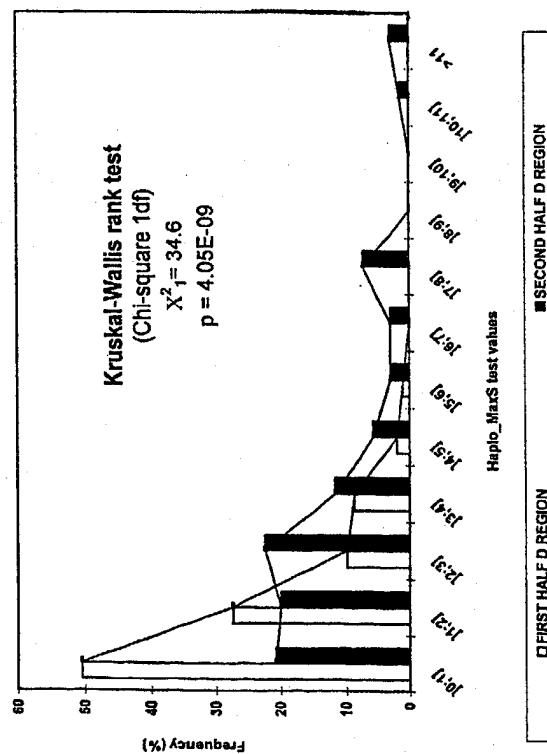


Figure 11A

Combination of 3 markers

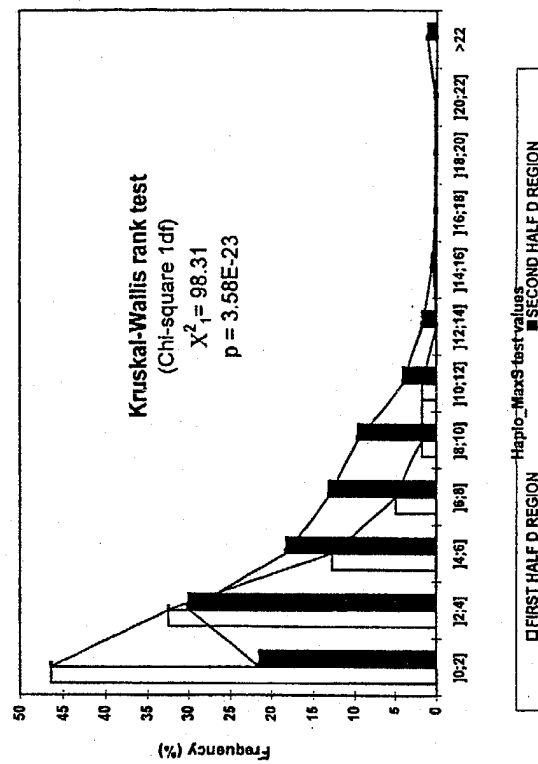


Figure 11B

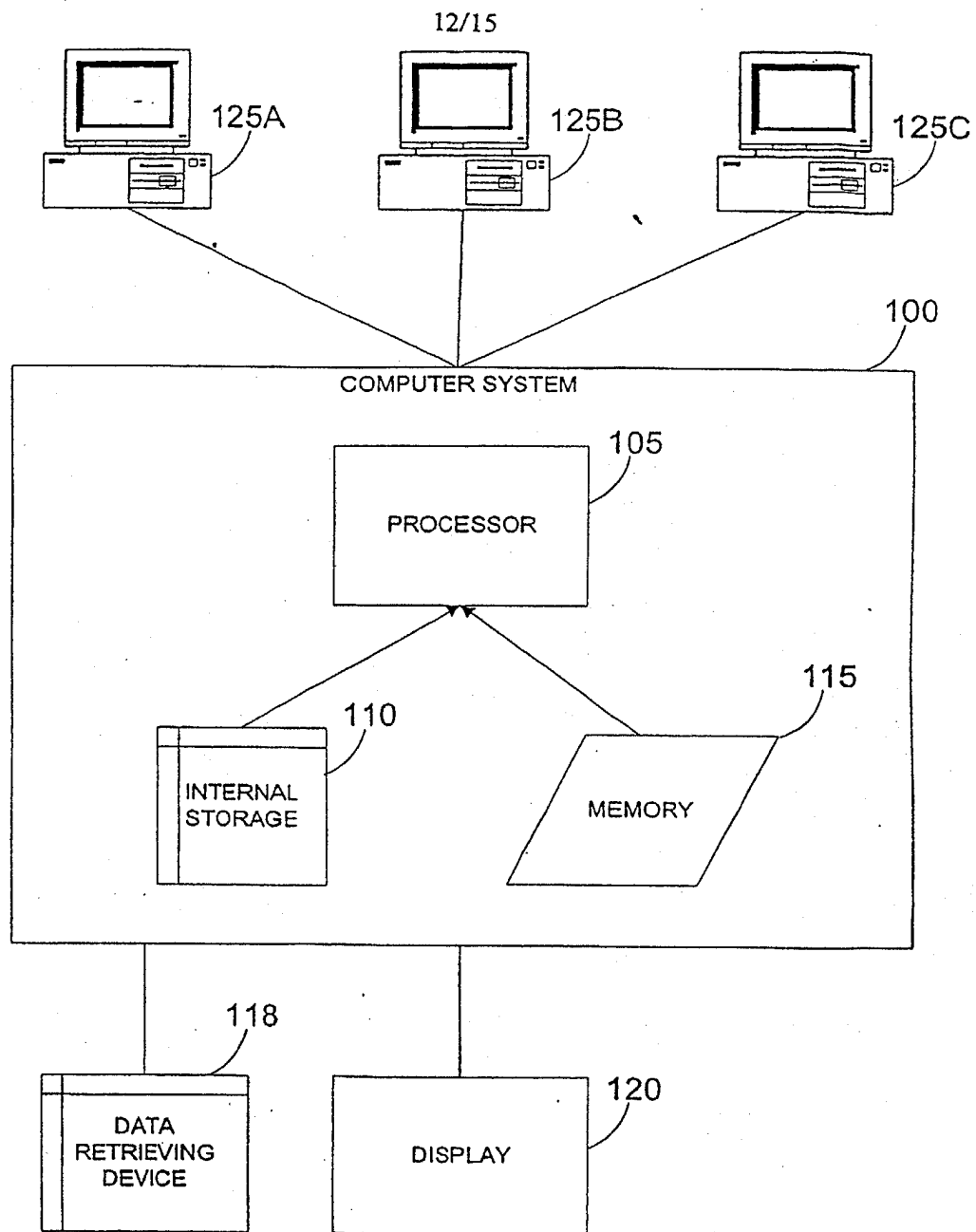


Figure 12

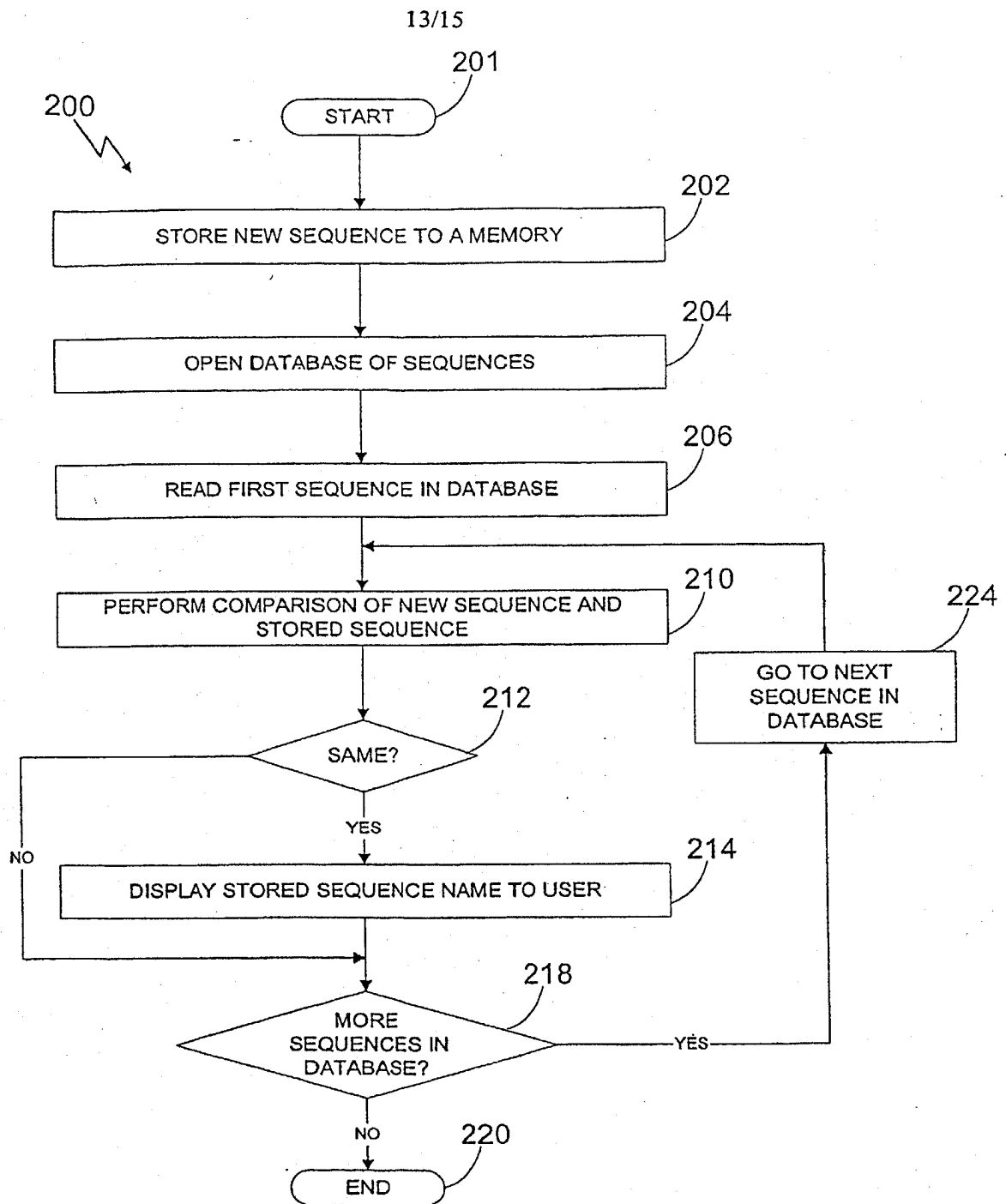


Figure 13

14/15

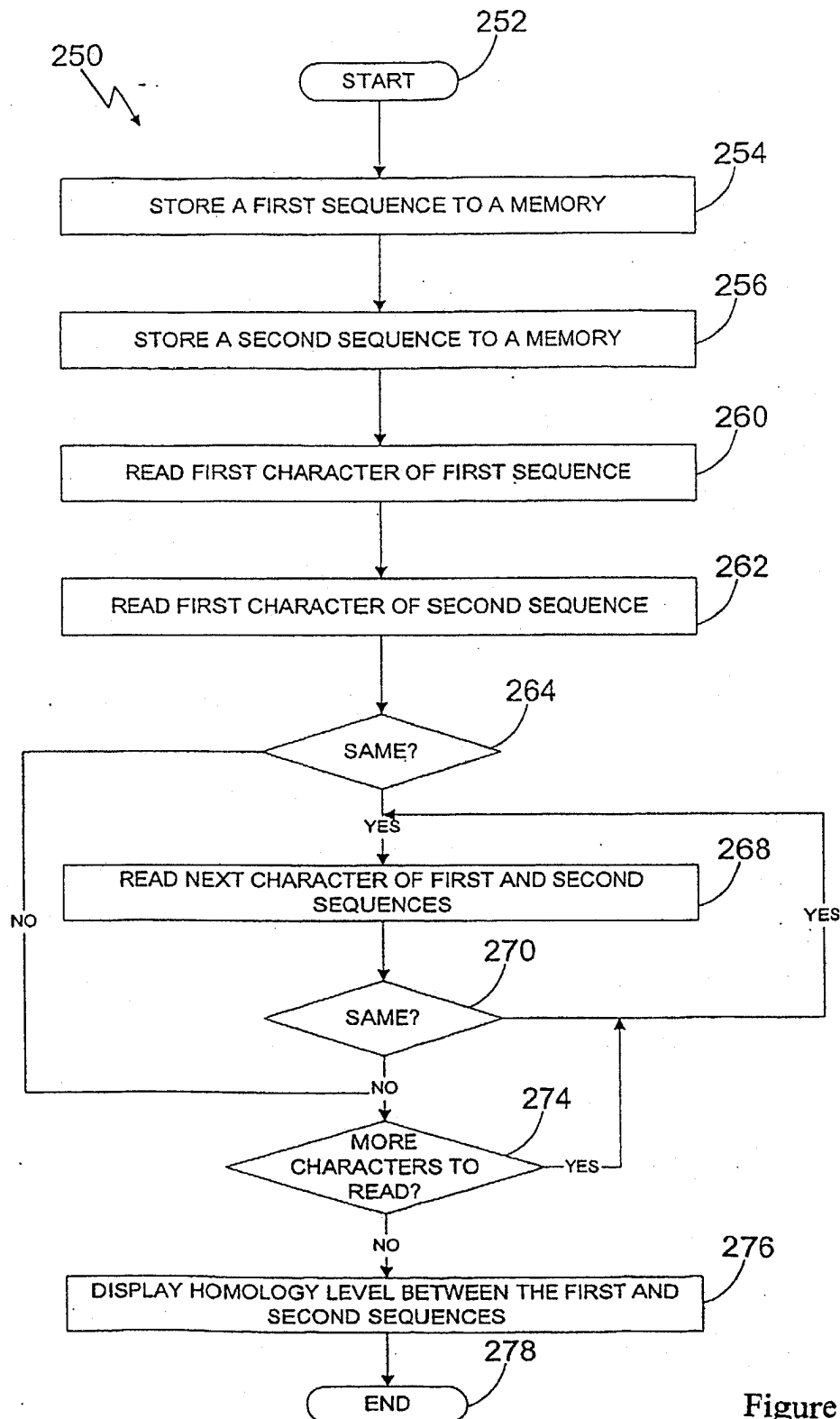


Figure 14

15/15

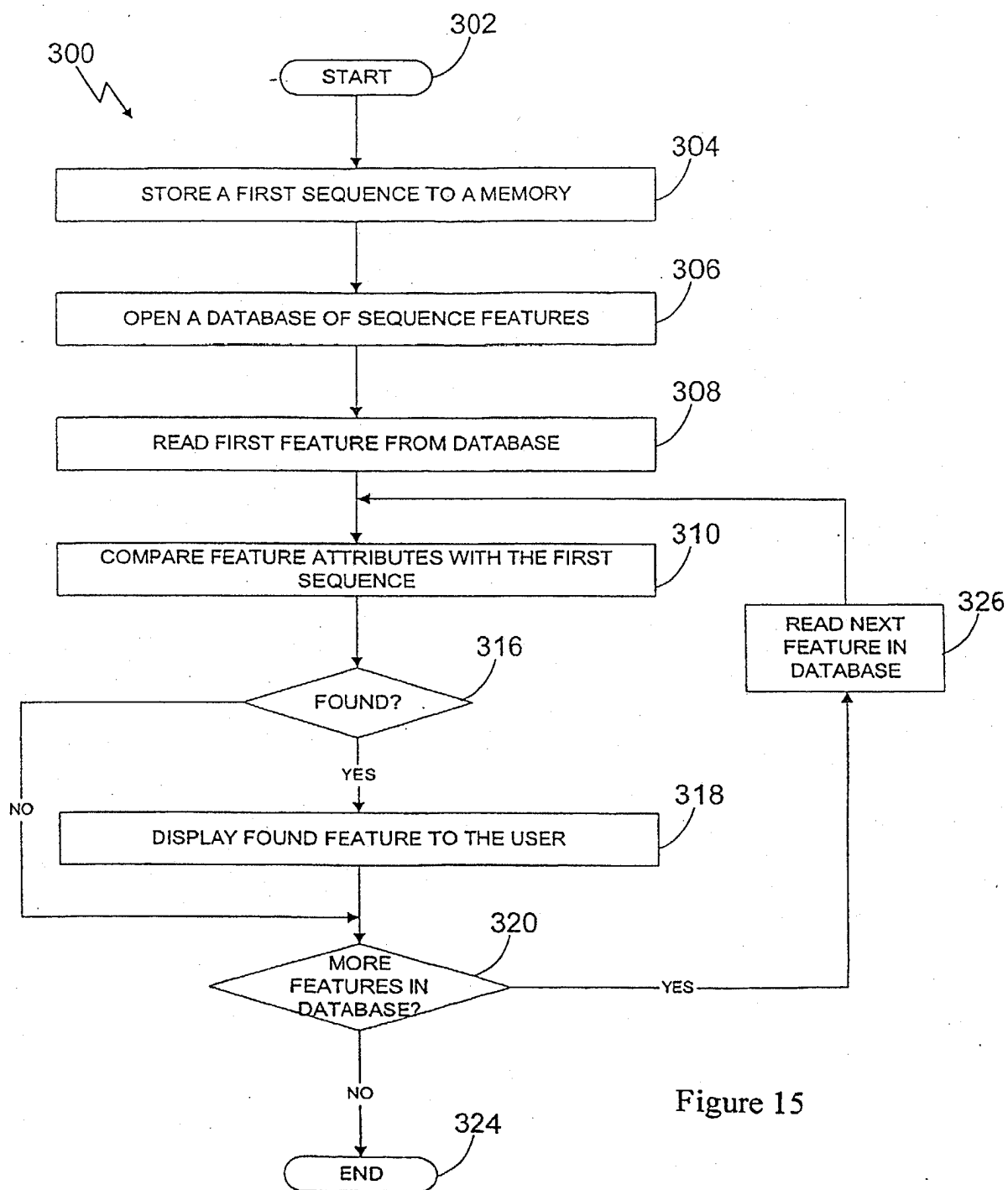


Figure 15

## SEQUENCE LISTING

&lt;110&gt; GENSET

&lt;120&gt; SCHIZOPHRENIA ASSOCIATED GENE, PROTEINS AND BIALLELIC MARKERS

&lt;130&gt; 53.WO2

&lt;150&gt; 60/168,088

&lt;151&gt; 1999-11-30

&lt;160&gt; 134

&lt;170&gt; Patent.pm

&lt;210&gt; 1

&lt;211&gt; 319608

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; 199122..201122

&lt;223&gt; 5'regulatory region

&lt;220&gt;

&lt;221&gt; exon

&lt;222&gt; 201123..201234

&lt;223&gt; exon S

&lt;220&gt;

&lt;221&gt; exon

&lt;222&gt; 201123..201560

&lt;223&gt; exon S2

&lt;220&gt;

&lt;221&gt; exon

&lt;222&gt; 214676..214793

&lt;223&gt; exon T

&lt;220&gt;

&lt;221&gt; exon

&lt;222&gt; 215702..215746

&lt;223&gt; exon U

&lt;220&gt;

&lt;221&gt; exon

&lt;222&gt; 216836..216994

&lt;223&gt; exon V

&lt;220&gt;

&lt;221&gt; exon

&lt;222&gt; 216836..217077

&lt;223&gt; exon V2

&lt;220&gt;

&lt;221&gt; exon



<222> 217671..217764  
<223> exon V1

<220>  
<221> exon  
<222> 227655..227736  
<223> exon V4

<220>  
<221> exon  
<222> 238715..238919  
<223> exon V3

<220>  
<221> exon  
<222> 240440..240673  
<223> exon W

<220>  
<221> exon  
<222> 240440..241153  
<223> exon W2

<220>  
<221> exon  
<222> 241072..241291  
<223> exon X

<220>  
<221> exon  
<222> 244353..244561  
<223> exon Y

<220>  
<221> exon  
<222> 246273..247802  
<223> exon Z

<220>  
<221> misc\_feature  
<222> 247803..249803  
<223> 3'regulatory region

<220>  
<221> allele  
<222> 8316  
<223> 99-27943-150 : polymorphic base G or C

<220>  
<221> allele  
<222> 21672  
<223> 99-27935-193 : polymorphic base G or C

<220>  
<221> allele  
<222> 65485  
<223> 8-128-33 : polymorphic base C or T

<220>  
<221> allele  
<222> 95396

<223> 99-31960-363 : polymorphic base A or G

<220>

<221> allele

<222> 107281

<223> 99-24656-260 : polymorphic base A or G

<220>

<221> allele

<222> 160640

<223> 99-24639-163 : polymorphic base A or C

<220>

<221> allele

<222> 160876

<223> 99-24634-108 : polymorphic base A or T

<220>

<221> allele

<222> 168974

<223> 99-7652-162 : polymorphic base A or G

<220>

<221> allele

<222> 170810

<223> 99-16100-147 : polymorphic base A or G

<220>

<221> allele

<222> 173358

<223> 99-5862-167 : polymorphic base A or G

<220>

<221> allele

<222> 189957

<223> 99-5919-215 : polymorphic base A or G

<220>

<221> allele

<222> 197163

<223> 99-24658-410 : polymorphic base A or G

<220>

<221> allele

<222> 200778

<223> 8-303-235 : polymorphic base A or G

<220>

<221> allele

<222> 202651

<223> 8-300-221 : polymorphic base A or G

<220>

<221> allele

<222> 202679

<223> 8-300-193 : polymorphic base A or G

<220>

<221> allele

<222> 203378

<223> 8-299-128 : polymorphic base A or T

<220>  
<221> allele  
<222> 204138  
<223> 8-296-213 : polymorphic base A or T

<220>  
<221> allele  
<222> 204605  
<223> 8-252-190 : polymorphic base C or T

<220>  
<221> allele  
<222> 204934  
<223> 99-24644-194 : polymorphic base A or G

<220>  
<221> allele  
<222> 205206  
<223> 8-295-248 : polymorphic base A or C

<220>  
<221> allele  
<222> 205329  
<223> 8-295-125 : polymorphic base C or T

<220>  
<221> allele  
<222> 206064  
<223> 8-293-130 : polymorphic base A or G

<220>  
<221> allele  
<222> 206545  
<223> 8-292-198 : polymorphic base A or G

<220>  
<221> allele  
<222> 207313  
<223> 8-251-322 : polymorphic base A or G

<220>  
<221> allele  
<222> 208285  
<223> 8-289-322 : polymorphic base A or G

<220>  
<221> allele  
<222> 208960  
<223> 8-287-249 : polymorphic base C or T

<220>  
<221> allele  
<222> 209123  
<223> 8-287-86 : polymorphic base A or T

<220>  
<221> allele  
<222> 209631  
<223> 8-285-319 : polymorphic base A or G

<220>  
<221> allele  
<222> 210361  
<223> 8-283-278 : polymorphic base G or C

<220>  
<221> allele  
<222> 210463  
<223> 8-283-176 : polymorphic base A or G

<220>  
<221> allele  
<222> 210486  
<223> 8-283-153 : polymorphic base G or C

<220>  
<221> allele  
<222> 210583  
<223> 8-283-56 : polymorphic base C or T

<220>  
<221> allele  
<222> 210879  
<223> 8-282-345 : polymorphic base G or C

<220>  
<221> allele  
<222> 210964  
<223> 8-282-260 : polymorphic base G or T

<220>  
<221> allele  
<222> 210979  
<223> 8-282-245 : polymorphic base A or C

<220>  
<221> allele  
<222> 211050  
<223> 8-282-174 : variable motif AAAGG or GAAGGAAGGAAGGAAGGAAGA

<220>  
<221> allele  
<222> 211132  
<223> 8-282-92 : polymorphic base A or T

<220>  
<221> allele  
<222> 211247  
<223> 8-281-367 : polymorphic base A or G

<220>  
<221> allele  
<222> 211315  
<223> 8-281-299 : polymorphic base A or G

<220>  
<221> allele  
<222> 211366  
<223> 8-281-248 : polymorphic base G or C

<220>

<221> allele  
<222> 212520  
<223> 8-279-197 : polymorphic base A or C

<220>  
<221> allele  
<222> 212821  
<223> 8-278-289 : polymorphic base C or T

<220>  
<221> allele  
<222> 213833  
<223> 8-277-250 : polymorphic base C or T

<220>  
<221> allele  
<222> 214857  
<223> 8-124-294 : polymorphic base A or G

<220>  
<221> allele  
<222> 215347  
<223> 8-273-383 : polymorphic base A or G

<220>  
<221> allele  
<222> 216967  
<223> 8-126-286 or 8-130-143 : polymorphic base A or G

<220>  
<221> allele  
<222> 219540  
<223> 99-5897-143 : polymorphic base A or C

<220>  
<221> allele  
<222> 220836  
<223> 99-24649-186 : polymorphic base A or G

<220>  
<221> allele  
<222> 223230  
<223> 8-247-337 : polymorphic base A or G

<220>  
<221> allele  
<222> 224213  
<223> 8-246-393 : polymorphic base A or T

<220>  
<221> allele  
<222> 224792  
<223> 8-245-270 : polymorphic base A or G

<220>  
<221> allele  
<222> 230395  
<223> 8-119-247 : polymorphic base A or G

<220>  
<221> allele

<222> 231677  
<223> 99-15870-400 : polymorphic base A or G

<220>  
<221> allele  
<222> 235315  
<223> 99-16087-219 : polymorphic base G or C

<220>  
<221> allele  
<222> 238789  
<223> 8-155-258 : polymorphic base C or T

<220>  
<221> allele  
<222> 239763  
<223> 8-136-267 or 99-16038-118 : polymorphic base C or T

<220>  
<221> allele  
<222> 241861  
<223> 99-15880-162 : polymorphic base A or G

<220>  
<221> allele  
<222> 243700  
<223> 8-271-227 : polymorphic base A or G

<220>  
<221> allele  
<222> 245264  
<223> 8-267-273 : deletion of GTGGGC

<220>  
<221> allele  
<222> 245663  
<223> 8-266-131 : polymorphic base A or C

<220>  
<221> allele  
<222> 245966  
<223> 8-265-208 : polymorphic base A or G

<220>  
<221> allele  
<222> 245971  
<223> 8-265-203 : polymorphic base G or C

<220>  
<221> allele  
<222> 247864  
<223> 99-25940-182 : polymorphic base C or T

<220>  
<221> allele  
<222> 248315  
<223> 99-16032-292 : polymorphic base G or T

<220>  
<221> allele  
<222> 255848

<223> 99-16105-152 : polymorphic base A or G

<220>

<221> allele

<222> 260099

<223> 99-16033-244 : polymorphic base A or G

<220>

<221> allele

<222> 279789

<223> 99-15875-165 : polymorphic base C or T

<220>

<221> allele

<222> 300365

<223> 99-16052-214 : polymorphic base A or G

<220>

<221> allele

<222> 312030

<223> 99-16047-115 : polymorphic base A or G

<220>

<221> allele

<222> 316014

<223> 99-25993-367 : polymorphic base A or G

<220>

<221> primer\_bind

<222> 7938..7958

<223> 99-27943.rp

<220>

<221> primer\_bind

<222> 8446..8465

<223> 99-27943.pu complement

<220>

<221> primer\_bind

<222> 21365..21385

<223> 99-27935.rp

<220>

<221> primer\_bind

<222> 21845..21864

<223> 99-27935.pu complement

<220>

<221> primer\_bind

<222> 65453..65471

<223> 8-128.pu

<220>

<221> primer\_bind

<222> 65856..65874

<223> 8-128.rp complement

<220>

<221> primer\_bind

<222> 95034..95053

<223> 99-31960.pu

<220>  
<221> primer\_bind  
<222> 95543..95563  
<223> 99-31960.rp complement

<220>  
<221> primer\_bind  
<222> 107022..107040  
<223> 99-24656.pu

<220>  
<221> primer\_bind  
<222> 107495..107513  
<223> 99-24656.rp complement

<220>  
<221> primer\_bind  
<222> 160279..160298  
<223> 99-24639.rp

<220>  
<221> primer\_bind  
<222> 160770..160787  
<223> 99-24634.pu

<220>  
<221> primer\_bind  
<222> 160785..160802  
<223> 99-24639.pu complement

<220>  
<221> primer\_bind  
<222> 161240..161257  
<223> 99-24634.rp complement

<220>  
<221> primer\_bind  
<222> 168813..168830  
<223> 99-7652.pu

<220>  
<221> primer\_bind  
<222> 169331..169351  
<223> 99-7652.rp complement

<220>  
<221> primer\_bind  
<222> 170666..170686  
<223> 99-16100.pu

<220>  
<221> primer\_bind  
<222> 171153..171173  
<223> 99-16100.rp complement

<220>  
<221> primer\_bind  
<222> 173065..173085  
<223> 99-5862.rp



<220>  
<221> primer\_bind  
<222> 173495..173514  
<223> 99-5862.pu complement

<220>  
<221> primer\_bind  
<222> 189753..189771  
<223> 99-5919.pu

<220>  
<221> primer\_bind  
<222> 190187..190207  
<223> 99-5919.rp complement

<220>  
<221> primer\_bind  
<222> 197116..197135  
<223> 99-24658.rp

<220>  
<221> primer\_bind  
<222> 197555..197572  
<223> 99-24658.pu complement

<220>  
<221> primer\_bind  
<222> 200587..200606  
<223> 8-303.rp

<220>  
<221> primer\_bind  
<222> 200994..201012  
<223> 8-303.pu complement

<220>  
<221> primer\_bind  
<222> 202440..202459  
<223> 8-300.rp

<220>  
<221> primer\_bind  
<222> 202852..202871  
<223> 8-300.pu complement

<220>  
<221> primer\_bind  
<222> 203061..203080  
<223> 8-299.rp

<220>  
<221> primer\_bind  
<222> 203486..203505  
<223> 8-299.pu complement

<220>  
<221> primer\_bind  
<222> 203943..203962  
<223> 8-296.rp

<220>

<221> primer\_bind  
<222> 204345..204362  
<223> 8-296.pu complement

<220>  
<221> primer\_bind  
<222> 204373..204392  
<223> 8-252.rp

<220>  
<221> primer\_bind  
<222> 204741..204758  
<223> 99-24644.pu

<220>  
<221> primer\_bind  
<222> 204777..204794  
<223> 8-252.pu complement

<220>  
<221> primer\_bind  
<222> 205034..205052  
<223> 8-295.rp

<220>  
<221> primer\_bind  
<222> 205222..205240  
<223> 99-24644.rp complement

<220>  
<221> primer\_bind  
<222> 205435..205453  
<223> 8-295.pu complement

<220>  
<221> primer\_bind  
<222> 205768..205786  
<223> 8-293.rp

<220>  
<221> primer\_bind  
<222> 206175..206193  
<223> 8-293.pu complement

<220>  
<221> primer\_bind  
<222> 206316..206333  
<223> 8-292.rp

<220>  
<221> primer\_bind  
<222> 206725..206742  
<223> 8-292.pu complement

<220>  
<221> primer\_bind  
<222> 207186..207205  
<223> 8-251.rp

<220>  
<221> primer\_bind

<222> 207616..207634  
<223> 8-251.pu complement

<220>  
<221> primer\_bind  
<222> 208186..208205  
<223> 8-289.rp

<220>  
<221> primer\_bind  
<222> 208588..208606  
<223> 8-289.pu complement

<220>  
<221> primer\_bind  
<222> 208786..208805  
<223> 8-287.rp

<220>  
<221> primer\_bind  
<222> 209189..209208  
<223> 8-287.pu complement

<220>  
<221> primer\_bind  
<222> 209520..209539  
<223> 8-285.rp

<220>  
<221> primer\_bind  
<222> 209931..209949  
<223> 8-285.pu complement

<220>  
<221> primer\_bind  
<222> 210219..210237  
<223> 8-283.rp

<220>  
<221> primer\_bind  
<222> 210620..210638  
<223> 8-283.pu complement

<220>  
<221> primer\_bind  
<222> 210801..210820  
<223> 8-282.rp

<220>  
<221> primer\_bind  
<222> 211182..211199  
<223> 8-281.rp

<220>  
<221> primer\_bind  
<222> 211205..211223  
<223> 8-282.pu complement

<220>  
<221> primer\_bind  
<222> 211596..211613

<223> 8-281.pu complement

<220>

<221> primer\_bind

<222> 212290..212309

<223> 8-279.rp

<220>

<221> primer\_bind

<222> 212662..212679

<223> 8-278.rp

<220>

<221> primer\_bind

<222> 212698..212716

<223> 8-279.pu complement

<220>

<221> primer\_bind

<222> 213091..213109

<223> 8-278.pu complement

<220>

<221> primer\_bind

<222> 213783..213803

<223> 8-277.rp

<220>

<221> primer\_bind

<222> 214062..214082

<223> 8-277.pu complement

<220>

<221> primer\_bind

<222> 214564..214581

<223> 8-124.pu

<220>

<221> primer\_bind

<222> 214965..214983

<223> 8-124.rp complement

<220>

<221> primer\_bind

<222> 215310..215328

<223> 8-273.rp

<220>

<221> primer\_bind

<222> 215711..215729

<223> 8-273.pu complement

<220>

<221> primer\_bind

<222> 216683..216702

<223> 8-130.rp

<220>

<221> primer\_bind

<222> 216687..216704

<223> 8-126.pu

<220>  
<221> primer\_bind  
<222> 217082..217101  
<223> 8-126.rp complement

<220>  
<221> primer\_bind  
<222> 217091..217109  
<223> 8-130.pu complement

<220>  
<221> primer\_bind  
<222> 219408..219425  
<223> 99-5897.pu

<220>  
<221> primer\_bind  
<222> 219882..219899  
<223> 99-5897.rp complement

<220>  
<221> primer\_bind  
<222> 220505..220522  
<223> 99-24649.rp

<220>  
<221> primer\_bind  
<222> 221004..221021  
<223> 99-24649.pu complement

<220>  
<221> primer\_bind  
<222> 223011..223028  
<223> 8-247.rp

<220>  
<221> primer\_bind  
<222> 223549..223566  
<223> 8-247.pu complement

<220>  
<221> primer\_bind  
<222> 223962..223978  
<223> 8-246.rp

<220>  
<221> primer\_bind  
<222> 224564..224581  
<223> 8-245.rp

<220>  
<221> primer\_bind  
<222> 224588..224605  
<223> 8-246.pu complement

<220>  
<221> primer\_bind  
<222> 225044..225061  
<223> 8-245.pu complement

<220>  
<221> primer\_bind  
<222> 230210..230227  
<223> 8-119.rp

<220>  
<221> primer\_bind  
<222> 230622..230641  
<223> 8-119.pu complement

<220>  
<221> primer\_bind  
<222> 231278..231298  
<223> 99-15870.pu

<220>  
<221> primer\_bind  
<222> 231729..231747  
<223> 99-15870.rp complement

<220>  
<221> primer\_bind  
<222> 235081..235101  
<223> 99-16087.rp

<220>  
<221> primer\_bind  
<222> 235515..235533  
<223> 99-16087.pu complement

<220>  
<221> primer\_bind  
<222> 238607..238626  
<223> 8-155.rp

<220>  
<221> primer\_bind  
<222> 239029..239046  
<223> 8-155.pu complement

<220>  
<221> primer\_bind  
<222> 239405..239425  
<223> 99-16038.rp

<220>  
<221> primer\_bind  
<222> 239606..239624  
<223> 8-136.rp

<220>  
<221> primer\_bind  
<222> 239862..239880  
<223> 99-16038.pu complement

<220>  
<221> primer\_bind  
<222> 240012..240029  
<223> 8-136.pu complement

<220>

<221> primer\_bind  
<222> 241700..241717  
<223> 99-15880.pu

<220>  
<221> primer\_bind  
<222> 242151..242171  
<223> 99-15880.rp complement

<220>  
<221> primer\_bind  
<222> 243484..243502  
<223> 8-271.rp

<220>  
<221> primer\_bind  
<222> 243909..243926  
<223> 8-271.pu complement

<220>  
<221> primer\_bind  
<222> 245105..245123  
<223> 8-267.rp

<220>  
<221> primer\_bind  
<222> 245372..245389  
<223> 8-266.rp

<220>  
<221> primer\_bind  
<222> 245518..245535  
<223> 8-267.pu complement

<220>  
<221> primer\_bind  
<222> 245754..245772  
<223> 8-265.rp

<220>  
<221> primer\_bind  
<222> 245775..245793  
<223> 8-266.pu complement

<220>  
<221> primer\_bind  
<222> 246155..246173  
<223> 8-265.pu complement

<220>  
<221> primer\_bind  
<222> 247513..247533  
<223> 99-25940.rp

<220>  
<221> primer\_bind  
<222> 248023..248043  
<223> 99-25940.pu complement

<220>  
<221> primer\_bind

<222> 248204..248223

<223> 99-16032.rp

<220>

<221> primer\_bind

<222> 248588..248606

<223> 99-16032.pu complement

<220>

<221> primer\_bind

<222> 255697..255715

<223> 99-16105.pu

<220>

<221> primer\_bind

<222> 256133..256152

<223> 99-16105.rp complement

<220>

<221> primer\_bind

<222> 259885..259902

<223> 99-16033.rp

<220>

<221> primer\_bind

<222> 260324..260342

<223> 99-16033.pu complement

<220>

<221> primer\_bind

<222> 279626..279644

<223> 99-15875.pu

<220>

<221> primer\_bind

<222> 280154..280173

<223> 99-15875.rp complement

<220>

<221> primer\_bind

<222> 300153..300170

<223> 99-16052.pu

<220>

<221> primer\_bind

<222> 300660..300680

<223> 99-16052.rp complement

<220>

<221> primer\_bind

<222> 311615..311632

<223> 99-16047.rp

<220>

<221> primer\_bind

<222> 312126..312144

<223> 99-16047.pu complement

<220>

<221> primer\_bind

<222> 315649..315668



<223> 99-25993.pu

<220>  
<221> primer\_bind  
<222> 316129..316147  
<223> 99-25993.rp complement

<220>  
<221> primer\_bind  
<222> 8297..8315  
<223> 99-27943-150.mis

<220>  
<221> primer\_bind  
<222> 8317..8335  
<223> 99-27943-150.mis complement

<220>  
<221> primer\_bind  
<222> 21653..21671  
<223> 99-27935-193.mis

<220>  
<221> primer\_bind  
<222> 21673..21691  
<223> 99-27935-193.mis complement

<220>  
<221> primer\_bind  
<222> 65466..65484  
<223> 8-128-33.mis

<220>  
<221> primer\_bind  
<222> 65486..65504  
<223> 8-128-33.mis complement

<220>  
<221> primer\_bind  
<222> 95377..95395  
<223> 99-31960-363.mis

<220>  
<221> primer\_bind  
<222> 95397..95415  
<223> 99-31960-363.mis complement

<220>  
<221> primer\_bind  
<222> 107262..107280  
<223> 99-24656-260.mis

<220>  
<221> primer\_bind  
<222> 107282..107300  
<223> 99-24656-260.mis complement

<220>  
<221> primer\_bind  
<222> 160621..160639  
<223> 99-24639-163.mis

<220>  
<221> primer\_bind  
<222> 160641..160659  
<223> 99-24639-163.mis complement

<220>  
<221> primer\_bind  
<222> 160857..160875  
<223> 99-24634-108.mis

<220>  
<221> primer\_bind  
<222> 160877..160895  
<223> 99-24634-108.mis complement

<220>  
<221> primer\_bind  
<222> 168955..168973  
<223> 99-7652-162.mis

<220>  
<221> primer\_bind  
<222> 168975..168993  
<223> 99-7652-162.mis complement

<220>  
<221> primer\_bind  
<222> 170791..170809  
<223> 99-16100-147.mis

<220>  
<221> primer\_bind  
<222> 170811..170829  
<223> 99-16100-147.mis complement

<220>  
<221> primer\_bind  
<222> 173339..173357  
<223> 99-5862-167.mis

<220>  
<221> primer\_bind  
<222> 173359..173377  
<223> 99-5862-167.mis complement

<220>  
<221> primer\_bind  
<222> 189938..189956  
<223> 99-5919-215.mis

<220>  
<221> primer\_bind  
<222> 189958..189976  
<223> 99-5919-215.mis complement

<220>  
<221> primer\_bind  
<222> 197144..197162  
<223> 99-24658-410.mis

<220>  
<221> primer\_bind  
<222> 197164..197182  
<223> 99-24658-410.mis complement

<220>  
<221> primer\_bind  
<222> 200759..200777  
<223> 8-303-235.mis

<220>  
<221> primer\_bind  
<222> 200779..200797  
<223> 8-303-235.mis complement

<220>  
<221> primer\_bind  
<222> 202632..202650  
<223> 8-300-221.mis

<220>  
<221> primer\_bind  
<222> 202652..202670  
<223> 8-300-221.mis complement

<220>  
<221> primer\_bind  
<222> 202660..202678  
<223> 8-300-193.mis

<220>  
<221> primer\_bind  
<222> 202680..202698  
<223> 8-300-193.mis complement

<220>  
<221> primer\_bind  
<222> 203359..203377  
<223> 8-299-128.mis

<220>  
<221> primer\_bind  
<222> 203379..203397  
<223> 8-299-128.mis complement

<220>  
<221> primer\_bind  
<222> 204119..204137  
<223> 8-296-213.mis

<220>  
<221> primer\_bind  
<222> 204139..204157  
<223> 8-296-213.mis complement

<220>  
<221> primer\_bind  
<222> 204586..204604  
<223> 8-252-190.mis

<220>

<221> primer\_bind  
<222> 204606..204624  
<223> 8-252-190.mis complement

<220>  
<221> primer\_bind  
<222> 204915..204933 -  
<223> 99-24644-194.mis

<220>  
<221> primer\_bind  
<222> 204935..204953  
<223> 99-24644-194.mis complement

<220>  
<221> primer\_bind  
<222> 205187..205205  
<223> 8-295-248.mis

<220>  
<221> primer\_bind  
<222> 205207..205225  
<223> 8-295-248.mis complement

<220>  
<221> primer\_bind  
<222> 205310..205328  
<223> 8-295-125.mis

<220>  
<221> primer\_bind  
<222> 205330..205348  
<223> 8-295-125.mis complement

<220>  
<221> primer\_bind  
<222> 206045..206063  
<223> 8-293-130.mis

<220>  
<221> primer\_bind  
<222> 206065..206083  
<223> 8-293-130.mis complement

<220>  
<221> primer\_bind  
<222> 206526..206544  
<223> 8-292-198.mis

<220>  
<221> primer\_bind  
<222> 206546..206564  
<223> 8-292-198.mis complement

<220>  
<221> primer\_bind  
<222> 207294..207312  
<223> 8-251-322.mis

<220>  
<221> primer\_bind

<222> 207314..207332  
<223> 8-251-322.mis complement

<220>  
<221> primer\_bind  
<222> 208266..208284  
<223> 8-289-322.mis

<220>  
<221> primer\_bind  
<222> 208286..208304  
<223> 8-289-322.mis complement

<220>  
<221> primer\_bind  
<222> 208941..208959  
<223> 8-287-249.mis

<220>  
<221> primer\_bind  
<222> 208961..208979  
<223> 8-287-249.mis complement

<220>  
<221> primer\_bind  
<222> 209104..209122  
<223> 8-287-86.mis

<220>  
<221> primer\_bind  
<222> 209124..209142  
<223> 8-287-86.mis complement

<220>  
<221> primer\_bind  
<222> 209612..209630  
<223> 8-285-319.mis

<220>  
<221> primer\_bind  
<222> 209632..209650  
<223> 8-285-319.mis complement

<220>  
<221> primer\_bind  
<222> 210342..210360  
<223> 8-283-278.mis

<220>  
<221> primer\_bind  
<222> 210362..210380  
<223> 8-283-278.mis complement

<220>  
<221> primer\_bind  
<222> 210444..210462  
<223> 8-283-176.mis

<220>  
<221> primer\_bind  
<222> 210464..210482

<223> 8-283-176.mis complement

<220>  
<221> primer\_bind  
<222> 210467..210485  
<223> 8-283-153.mis

<220>  
<221> primer\_bind  
<222> 210487..210505  
<223> 8-283-153.mis complement

<220>  
<221> primer\_bind  
<222> 210564..210582  
<223> 8-283-56.mis

<220>  
<221> primer\_bind  
<222> 210584..210602  
<223> 8-283-56.mis complement

<220>  
<221> primer\_bind  
<222> 210860..210878  
<223> 8-282-345.mis

<220>  
<221> primer\_bind  
<222> 210880..210898  
<223> 8-282-345.mis complement

<220>  
<221> primer\_bind  
<222> 210945..210963  
<223> 8-282-260.mis

<220>  
<221> primer\_bind  
<222> 210960..210978  
<223> 8-282-245.mis

<220>  
<221> primer\_bind  
<222> 210965..210983  
<223> 8-282-260.mis complement

<220>  
<221> primer\_bind  
<222> 210980..210998  
<223> 8-282-245.mis complement

<220>  
<221> primer\_bind  
<222> 211113..211131  
<223> 8-282-92.mis

<220>  
<221> primer\_bind  
<222> 211133..211151  
<223> 8-282-92.mis complement

<220>  
<221> primer\_bind  
<222> 211228..211246  
<223> 8-281-367.mis

<220>  
<221> primer\_bind  
<222> 211248..211266  
<223> 8-281-367.mis complement

<220>  
<221> primer\_bind  
<222> 211296..211314  
<223> 8-281-299.mis

<220>  
<221> primer\_bind  
<222> 211316..211334  
<223> 8-281-299.mis complement

<220>  
<221> primer\_bind  
<222> 211347..211365  
<223> 8-281-248.mis

<220>  
<221> primer\_bind  
<222> 211367..211385  
<223> 8-281-248.mis complement

<220>  
<221> primer\_bind  
<222> 212501..212519  
<223> 8-279-197.mis

<220>  
<221> primer\_bind  
<222> 212521..212539  
<223> 8-279-197.mis complement

<220>  
<221> primer\_bind  
<222> 212802..212820  
<223> 8-278-289.mis

<220>  
<221> primer\_bind  
<222> 212822..212840  
<223> 8-278-289.mis complement

<220>  
<221> primer\_bind  
<222> 213814..213832  
<223> 8-277-250.mis

<220>  
<221> primer\_bind  
<222> 213834..213852  
<223> 8-277-250.mis complement

<220>  
<221> primer\_bind  
<222> 214838..214856  
<223> 8-124-294.mis

<220>  
<221> primer\_bind  
<222> 214858..214876  
<223> 8-124-294.mis complement

<220>  
<221> primer\_bind  
<222> 215328..215346  
<223> 8-273-383.mis

<220>  
<221> primer\_bind  
<222> 215348..215366  
<223> 8-273-383.mis complement

<220>  
<221> primer\_bind  
<222> 216948..216966  
<223> 8-126-286.mis

<220>  
<221> primer\_bind  
<222> 216948..216966  
<223> 8-130-143.mis

<220>  
<221> primer\_bind  
<222> 216968..216986  
<223> 8-126-286.mis complement

<220>  
<221> primer\_bind  
<222> 216968..216986  
<223> 8-130-143.mis complement

<220>  
<221> primer\_bind  
<222> 219521..219539  
<223> 99-5897-143.mis

<220>  
<221> primer\_bind  
<222> 219541..219559  
<223> 99-5897-143.mis complement

<220>  
<221> primer\_bind  
<222> 220817..220835  
<223> 99-24649-186.mis

<220>  
<221> primer\_bind  
<222> 220837..220855  
<223> 99-24649-186.mis complement

<220>



<221> primer\_bind  
<222> 223211..223229  
<223> 8-247-337.mis

<220>  
<221> primer\_bind  
<222> 223231..223249  
<223> 8-247-337.mis complement

<220>  
<221> primer\_bind  
<222> 224194..224212  
<223> 8-246-393.mis

<220>  
<221> primer\_bind  
<222> 224214..224232  
<223> 8-246-393.mis complement

<220>  
<221> primer\_bind  
<222> 224773..224791  
<223> 8-245-270.mis

<220>  
<221> primer\_bind  
<222> 224793..224811  
<223> 8-245-270.mis complement

<220>  
<221> primer\_bind  
<222> 230376..230394  
<223> 8-119-247.mis

<220>  
<221> primer\_bind  
<222> 230396..230414  
<223> 8-119-247.mis complement

<220>  
<221> primer\_bind  
<222> 231658..231676  
<223> 99-15870-400.mis

<220>  
<221> primer\_bind  
<222> 231678..231696  
<223> 99-15870-400.mis complement

<220>  
<221> primer\_bind  
<222> 235296..235314  
<223> 99-16087-219.mis

<220>  
<221> primer\_bind  
<222> 235316..235334  
<223> 99-16087-219.mis complement

<220>  
<221> primer\_bind

<222> 238770..238788

<223> 8-155-258.mis

<220>

<221> primer\_bind

<222> 238790..238808

<223> 8-155-258.mis complement

<220>

<221> primer\_bind

<222> 239744..239762

<223> 8-136-267.mis

<220>

<221> primer\_bind

<222> 239744..239762

<223> 99-16038-118.mis

<220>

<221> primer\_bind

<222> 239764..239782

<223> 8-136-267.mis complement

<220>

<221> primer\_bind

<222> 239764..239782

<223> 99-16038-118.mis complement

<220>

<221> primer\_bind

<222> 241842..241860

<223> 99-15880-162.mis

<220>

<221> primer\_bind

<222> 241862..241880

<223> 99-15880-162.mis complement

<220>

<221> primer\_bind

<222> 243681..243699

<223> 8-271-227.mis

<220>

<221> primer\_bind

<222> 243701..243719

<223> 8-271-227.mis complement

<220>

<221> primer\_bind

<222> 245644..245662

<223> 8-266-131.mis

<220>

<221> primer\_bind

<222> 245664..245682

<223> 8-266-131.mis complement

<220>

<221> primer\_bind

<222> 245947..245965

<223> 8-265-208.mis

<220>

<221> primer\_bind

<222> 245952..245970

<223> 8-265-203.mis

<220>

<221> primer\_bind

<222> 245967..245985

<223> 8-265-208.mis complement

<220>

<221> primer\_bind

<222> 245972..245990

<223> 8-265-203.mis complement

<220>

<221> primer\_bind

<222> 247845..247863

<223> 99-25940-182.mis

<220>

<221> primer\_bind

<222> 247865..247883

<223> 99-25940-182.mis complement

<220>

<221> primer\_bind

<222> 248296..248314

<223> 99-16032-292.mis

<220>

<221> primer\_bind

<222> 248316..248334

<223> 99-16032-292.mis complement

<220>

<221> primer\_bind

<222> 255829..255847

<223> 99-16105-152.mis

<220>

<221> primer\_bind

<222> 255849..255867

<223> 99-16105-152.mis complement

<220>

<221> primer\_bind

<222> 260080..260098

<223> 99-16033-244.mis

<220>

<221> primer\_bind

<222> 260100..260118

<223> 99-16033-244.mis complement

<220>

<221> primer\_bind

<222> 279770..279788

<223> 99-15875-165.mis

<220>  
<221> primer\_bind  
<222> 279790..279808  
<223> 99-15875-165.mis complement

<220>  
<221> primer\_bind  
<222> 300346..300364  
<223> 99-16052-214.mis

<220>  
<221> primer\_bind  
<222> 300366..300384  
<223> 99-16052-214.mis complement

<220>  
<221> primer\_bind  
<222> 312011..312029  
<223> 99-16047-115.mis

<220>  
<221> primer\_bind  
<222> 312031..312049  
<223> 99-16047-115.mis complement

<220>  
<221> primer\_bind  
<222> 315995..316013  
<223> 99-25993-367.mis

<220>  
<221> primer\_bind  
<222> 316015..316033  
<223> 99-25993-367.mis complement

<220>  
<221> misc\_binding  
<222> 8304..8328  
<223> 99-27943-150.probe

<220>  
<221> misc\_binding  
<222> 21660..21684  
<223> 99-27935-193.probe

<220>  
<221> misc\_binding  
<222> 65473..65497  
<223> 8-128-33.probe

<220>  
<221> misc\_binding  
<222> 95384..95408  
<223> 99-31960-363.probe

<220>  
<221> misc\_binding  
<222> 107269..107293  
<223> 99-24656-260.probe

<220>  
<221> misc\_binding  
<222> 160628..160652  
<223> 99-24639-163.probe

<220>  
<221> misc\_binding  
<222> 160864..160888  
<223> 99-24634-108.probe

<220>  
<221> misc\_binding  
<222> 168962..168986  
<223> 99-7652-162.probe

<220>  
<221> misc\_binding  
<222> 170798..170822  
<223> 99-16100-147.probe

<220>  
<221> misc\_binding  
<222> 173346..173370  
<223> 99-5862-167.probe

<220>  
<221> misc\_binding  
<222> 189945..189969  
<223> 99-5919-215.probe

<220>  
<221> misc\_binding  
<222> 197151..197175  
<223> 99-24658-410.probe

<220>  
<221> misc\_binding  
<222> 200766..200790  
<223> 8-303-235.probe

<220>  
<221> misc\_binding  
<222> 202639..202663  
<223> 8-300-221.probe

<220>  
<221> misc\_binding  
<222> 202667..202691  
<223> 8-300-193.probe

<220>  
<221> misc\_binding  
<222> 203366..203390  
<223> 8-299-128.probe

<220>  
<221> misc\_binding  
<222> 204126..204150  
<223> 8-296-213.probe

<220>

<221> misc\_binding  
<222> 204593..204617  
<223> 8-252-190.probe

<220>  
<221> misc\_binding  
<222> 204922..204946  
<223> 99-24644-194.probe

<220>  
<221> misc\_binding  
<222> 205194..205218  
<223> 8-295-248.probe

<220>  
<221> misc\_binding  
<222> 205317..205341  
<223> 8-295-125.probe

<220>  
<221> misc\_binding  
<222> 206052..206076  
<223> 8-293-130.probe

<220>  
<221> misc\_binding  
<222> 206533..206557  
<223> 8-292-198.probe

<220>  
<221> misc\_binding  
<222> 207301..207325  
<223> 8-251-322.probe

<220>  
<221> misc\_binding  
<222> 208273..208297  
<223> 8-289-322.probe

<220>  
<221> misc\_binding  
<222> 208948..208972  
<223> 8-287-249.probe

<220>  
<221> misc\_binding  
<222> 209111..209135  
<223> 8-287-86.probe

<220>  
<221> misc\_binding  
<222> 209619..209643  
<223> 8-285-319.probe

<220>  
<221> misc\_binding  
<222> 210349..210373  
<223> 8-283-278.probe

<220>  
<221> misc\_binding

<222> 210451..210475  
<223> 8-283-176.probe

<220>  
<221> misc\_binding  
<222> 210474..210498  
<223> 8-283-153.probe

<220>  
<221> misc\_binding  
<222> 210571..210595  
<223> 8-283-56.probe

<220>  
<221> misc\_binding  
<222> 210867..210891  
<223> 8-282-345.probe

<220>  
<221> misc\_binding  
<222> 210952..210976  
<223> 8-282-260.probe

<220>  
<221> misc\_binding  
<222> 210967..210991  
<223> 8-282-245.probe

<220>  
<221> misc\_binding  
<222> 211120..211144  
<223> 8-282-92.probe

<220>  
<221> misc\_binding  
<222> 211235..211259  
<223> 8-281-367.probe

<220>  
<221> misc\_binding  
<222> 211303..211327  
<223> 8-281-299.probe

<220>  
<221> misc\_binding  
<222> 211354..211378  
<223> 8-281-248.probe

<220>  
<221> misc\_binding  
<222> 212508..212532  
<223> 8-279-197.probe

<220>  
<221> misc\_binding  
<222> 212809..212833  
<223> 8-278-289.probe

<220>  
<221> misc\_binding  
<222> 213821..213845

<223> 8-277-250.probe

<220>

<221> misc\_binding

<222> 214845..214869

<223> 8-124-294.probe

<220>

<221> misc\_binding

<222> 215335..215359

<223> 8-273-383.probe

<220>

<221> misc\_binding

<222> 216955..216979

<223> 8-126-286.probe

<220>

<221> misc\_binding

<222> 216955..216979

<223> 8-130-143.probe

<220>

<221> misc\_binding

<222> 219528..219552

<223> 99-5897-143.probe

<220>

<221> misc\_binding

<222> 220824..220848

<223> 99-24649-186.probe

<220>

<221> misc\_binding

<222> 223218..223242

<223> 8-247-337.probe

<220>

<221> misc\_binding

<222> 224201..224225

<223> 8-246-393.probe

<220>

<221> misc\_binding

<222> 224780..224804

<223> 8-245-270.probe

<220>

<221> misc\_binding

<222> 230383..230407

<223> 8-119-247.probe

<220>

<221> misc\_binding

<222> 231665..231689

<223> 99-15870-400.probe

<220>

<221> misc\_binding

<222> 235303..235327

<223> 99-16087-219.probe



<220>  
<221> misc\_binding  
<222> 238777..238801  
<223> 8-155-258.probe

<220>  
<221> misc\_binding  
<222> 239751..239775  
<223> 8-136-267.probe

<220>  
<221> misc\_binding  
<222> 239751..239775  
<223> 99-16038-118.probe

<220>  
<221> misc\_binding  
<222> 241849..241873  
<223> 99-15880-162.probe

<220>  
<221> misc\_binding  
<222> 243688..243712  
<223> 8-271-227.probe

<220>  
<221> misc\_binding  
<222> 245651..245675  
<223> 8-266-131.probe

<220>  
<221> misc\_binding  
<222> 245954..245978  
<223> 8-265-208.probe

<220>  
<221> misc\_binding  
<222> 245959..245983  
<223> 8-265-203.probe

<220>  
<221> misc\_binding  
<222> 247852..247876  
<223> 99-25940-182.probe

<220>  
<221> misc\_binding  
<222> 248303..248327  
<223> 99-16032-292.probe

<220>  
<221> misc\_binding  
<222> 255836..255860  
<223> 99-16105-152.probe

<220>  
<221> misc\_binding  
<222> 260087..260111  
<223> 99-16033-244.probe

<220>  
 <221> misc\_binding  
 <222> 279777..279801  
 <223> 99-15875-165.probe

<220>  
 <221> misc\_binding  
 <222> 300353..300377  
 <223> 99-16052-214.probe

<220>  
 <221> misc\_binding  
 <222> 312018..312042  
 <223> 99-16047-115.probe

<220>  
 <221> misc\_binding  
 <222> 316002..316026  
 <223> 99-25993-367.probe

<220>  
 <221> misc\_feature  
 <222> 4222,11245,27205..27206,27212..27213,27240,27347,27378,28258  
 108818,128283,137062..137063,164757,172148..172150,180820,184929  
 194460,200104,207061..207062,207949,239085,249787,249806,250156  
 250189,250841,251114,264938,265019,265078,274640,277770,314308  
 318680,318731  
 <223> n=a, g, c or t

<400> 1  
 cgggcggaatt cgagctcggt acccggggat catctttgtc atagattgta gtaaaaataa 60  
 tctgtgaaaa gtgactgatt attgacaggt agccttgggt atgcagctca agtaggcatt 120  
 ataaaataaa aaacatcaca ttccatgtat tttcaaaaaa agcttctggc ttactcaaaa 180  
 ggtaagtgtt ctttaaaciaa cttcaatgga aaaggagttt cttccaaatg cttggagaac 240  
 tccttctcaa gtccggaata aagcaaggtc tacagagaca aagtataaat aaaaccatgg 300  
 tgctgagctg gtaccactat tgggctgcta agcactgcta agcattgcct tcagatgagt 360  
 tacctagcct ttgtaaagct cagattccac agctataaac tgagatgaat aataatacac 420  
 aaaacactaa gctgttgagt agattaaatg aaattattta tgtagagcaa atagaacagt 480  
 actgtacctt ataaaattta aaacaaaaat gtaaattctg atttttaccc aagggtcaac 540  
 tccgcacaat ttgtcaatgg acccttcccc ggccctccag aagctgaata tcgttgtgta 600  
 tcatatccaa attcagtatc accagggcac agagtaagtg ccaactgtga ggccatagct 660  
 atgctgcac tccctgaaa cacatttgtt accctcattc ttggagttga gtctttgcat 720  
 ttcatcctcc agagtggaga tagactgcgt gttggagagg tcagggaaat tatgtctttc 780  
 tatgtctagc cttttatagg tcgattattt gaatgcttga cagcatctgt ccacttatct 840  
 atccctgcaa taataaataa cgggactaaa gtactcgtta tctctagtgg gagagtgcac 900  
 atattttata cgtcaattct ttaaaagggg aaattctttc agagcttcag aaaatatcac 960  
 ggaagagtta aaagaatgtt tgagaagaac atgtctgagt attacctact agcattacaa 1020  
 ttttactgct ttaaagtaat tataattttc attgcaggat ctataaaaaa ggcttttacc 1080  
 tgacggtaat tcttcttgta cttacaggtc attattacaa tgaagggaag aaaggaagac 1140  
 atgaagggga aatgggacta tttaccaagt gtatctgcat agactaaagg agacaagagg 1200  
 aagtactgag caccttgatt cctgcagata attatttaag aaccataaga attctttcct 1260  
 agcttctcat aagcctgctc cttcaaaggg tgagtagttg acatgatttt acatccatt 1320  
 tatatttctt tattagataa ttttatttca aatctagtct tattttttct gaaccttctg 1380  
 ggatatataa cattttatat tatcatttat gtcataaga atgtagattt atattctttt 1440  
 ctattcacat gttttatgta ataattgaac acctctaaat gacagatact atcattttat 1500  
 atattgttaa taagtattct catggctggg cgtgggtggc caccgccgta atcccagcac 1560  
 tttgggaggg cgaggtggga ggatcacctg aggtcaggag tttgagatca gcctggcgaa 1620  
 cgtgggtgaa tcccatctct actaaaaaca acaaaaatta gctgggcatg gtagcatgcg 1680  
 cctatagtcc cagctactag ggaggccaag gcaggagaat cgatcgctca aaactgggag 1740  
 gcagaggttt cagtggagctg atatcgtgcc actgcactct ggccctgggca acagaatgag 1800  
 accctgcttc aataaataaa taaatttaatt aattaattaa taaagtatct taatgacaaa 1860

tatttaaata	tatataaata	tttacattta	catatgggaa	gagagatctg	aaattttaag	1920
caacttttat	ccaaaaacct	atttttttca	aagttcctcc	aacatctata	aaagaagtaa	1980
agtcacttat	agatctatct	tacttcaggg	cctatgtttt	cctctacccat	acaacactgc	2040
acattcacag	acacagaaat	cttttacatt	tctaaaataa	ataaaacatc	caaacttcag	2100
tttaaaaaaa	atggatctac	agatccacca	tagactaaac	gaccaaacat	agtcagaggt	2160
gtttacttgt	ttgtttcggt	cttaagggaa	tcttcaaact	accctagagc	ctcgtgtatc	2220
caatgagagc	agtctggaca	ttgaatgaat	gggtgatgtca	ttaaatgagc	tatggagcaa	2280
gatgtaagaa	cacgcttttag	agtaaataac	atattttggct	tctgaggtat	aaaaagggaa	2340
ggaggctttc	tcaggacatt	tgggtacctg	gtccttcata	ctcaggaggt	ttttccatct	2400
tcattgtcatg	gaataaaacta	tgcttttaaag	aaatgctgcg	tcagcagaaa	gtgtttcaaa	2460
tataaggaaa	ttgtcatagc	tttgaaatta	tgtcaaatat	catttttaact	aagagagcca	2520
ttgtctttcc	atatatttat	ctagaagaaa	atttactgga	caaaaataaat	atagaacaac	2580
ttcatgtgta	gtcattaatg	gtcggatagg	cagttactat	aaaatatgaa	taactctaaa	2640
atcatttttag	ccatttggtt	tttaataaatc	tcactttgtga	ttgattttatt	aaaaacaggg	2700
cctccagaag	aaagaaactt	aatacctttc	ctccatcatt	atgaaggcac	tagtactgac	2760
gtaactaaaa	ttagatggca	gcttgaggga	ggaagagatg	atgtagagat	ggtgtgtaat	2820
ctgtgtaaac	agtctctcaa	ctgtctagag	ttagctgtac	gaatctgttg	ttggctccag	2880
ttataagaac	caagatatgt	atgtaaggac	aaacatcctt	aatttgctta	gcaagtcgta	2940
tgtattttgat	agaccaccaa	ttttatttta	ttttatttta	tttttgagat	ggagtcttgc	3000
tctgtcacct	gggctggagt	gcagtgggcg	gatctcggct	cactgcaaac	tccatttagc	3060
aggttcaagt	gattcttctg	cctcagcctc	ccgagtagct	gggattacaa	gtgctcaaaa	3120
ccacccccgg	ttaatgtttt	tttcttttct	tttcttttct	tttcttttct	tttcttttct	3180
tttcttttct	tttcttttct	tttttttagta	gtagacacag	ggtttccacta	tgttggccag	3240
gctgggtctca	aactcctgac	ctcgtgatct	gccccactcg	gcctcccaaa	gtgctgggat	3300
tacgggatta	caagcgtgag	ccactgcacc	agccctgatt	tttaaaaggg	attttagaat	3360
ttcagttgaa	catgtgtttt	tgaagtaact	ccatattgaa	caggggaatc	agagtaagta	3420
aaccactaca	ggaaaatcag	gaaggcagac	tccaatttat	gttttctgtg	tataacctac	3480
tttaataaag	aattaaaaat	cataagagta	gttttatcta	gtggcctaga	acttgattat	3540
gaaaagtaag	acaaaggaaa	tctgacctca	tgccagtgtc	tggaaagcctt	acaggaactg	3600
aaaggaccca	cgtggaacag	ctggcaactg	aaactcatct	tgtcaatttc	aagtgtagtc	3660
cctcttccca	attcaccgga	gaaaacacag	ccagctttgc	atacagtgtc	tgggtgcacc	3720
agagaaaaca	cagtcagctt	tgcatacaat	atctgtgtgc	agaatgacca	aacgattggc	3780
ttctatcctt	caggattgtg	aacagagagc	caaggcatac	aaccagtatg	cccactccgt	3840
cagctgcgag	acagaggaga	taaagaatgc	tgaacattat	gggtcactgc	tttgaagtat	3900
tcagaggcca	aaatagttca	aaataattgt	gttaatcatt	aaaagtagaa	gaagtggaa	3960
ccaaaattat	gcctgcctat	atttcagttt	tgaatgtgtg	gagatcagaa	atttaaaaaa	4020
aagtttatga	gtttccctt	ttaaactttg	agaaaagacc	ataagaaccg	gtgttttagt	4080
aaatcatctt	caagagtaat	ttactcataa	aaagaaaaga	actcatgata	ccgtgtaaac	4140
aaaaggctaa	taaatgctgg	tgtaaataag	attaagggag	gacaaagatg	aaggaaagca	4200
ggacataatca	aaaaacaaat	anaataagaa	attttcagag	cttgaatata	caaacaggca	4260
catcaaaaaga	ggttcaatgg	aaacctaatg	caatgaatgg	gacaattaat	gcttttaaac	4320
atcacttgga	accacagaaa	ataccaacag	aaaaatctgg	tcacgtacaa	gagataagag	4380
tcagggttcc	atggatgttc	ttcttagatg	agaaggacac	ttcactttta	tggcattgtt	4440
cccacaaatc	cataacccaa	gtttcatcat	gagcaagcct	cagccaaaga	aaaattgaga	4500
tgtattttac	aacatacccg	ttcaggccct	ttaaaaatgt	caaagttggc	caggcaccat	4560
ggcttccacc	tgtaatccca	gcactttggg	aggctgagat	gggtggatca	tttgaggtoa	4620
ggagtctcag	accagcctgg	acaatatggg	gaaaccccaa	ctgtactaaa	aaaagataac	4680
aaaattagcc	aggtgtgggtg	gtgtgggcct	gtaatcccag	ctacttgaga	ggctgaggca	4740
ggagaatcgc	ttgaaccocag	gagccaggag	gtggagggtg	cagttagcca	agatcatgcc	4800
actgtattcc	aacctgggcg	acagagcaaa	actccatctc	cccccaaaaa	aagcaaagtc	4860
gtgaaaaaca	aagtaagact	aaggaactgt	cagattagag	gggactgagg	aaacatgaaa	4920
attagatgaa	atagggtata	cagtttagaa	ttatagaaca	gaaagcagga	catcagttga	4980
acaactaaag	aaatacaaat	aaagtctatt	ttagttaata	ttgttgtacc	aataactaatt	5040
tcttagttgt	aacaaataaa	atatcattat	gtaagatggc	aacttcagga	aaagttgggg	5100
gaaagatata	caggaataact	ttgtattatc	ttagaaacta	ttatgtcaat	attttaatat	5160
aaaattatta	gaaattgtgt	cagaaacact	ggatgtttaga	agacaatgga	ggaatgcctc	5220
aaaaaatcta	cagagaactt	cttttcaatc	tatcatcata	gcataaatca	agtatgagct	5280
tcaaattatg	atattttata	atattgagaat	aattctgaaa	atttacctta	taccaatatt	5340
ttcttagaag	attacttttag	gatgtgctat	agtcaaacag	aggaataaac	taagaaaaag	5400
tatgatgtga	gatacagaaa	aatatataga	aacaatttag	aaaaataata	aaaagtagtc	5460
ctatgataat	agcaatataa	gaatgttttag	attggaatag	aagatgaatg	ccctcaagag	5520

tataacctct	gaaagaaaaa	ttaaataaat	aaaaaaattc	tgacaggata	cctgatcaga	5580
tagagatgtg	acaaaccagt	atgttgaaa	gtgatgacag	aaagagagaa	cggaagaga	5640
gaggcaggaa	ggaaaggga	aaaaaagaaa	gaaatgaaa	gcgatttctca	caccagttag	5700
aattggcaatt	attaaaaagt	caggaaacaa	caggtgctag	acaggatgtg	gagaaatagg	5760
aacactttta	cactattggt	gggactgtaa	actagtcca	ccattgtgga	agtcagtgtg	5820
gcgattcctc	agggatctag	aactagaaat	accatttgac	ccagtcaccc	cattactggc	5880
tatatacca	aaggattata	aatcatgctt	ctataaagac	acatgcacac	gtatgtttat	5940
tgcggcacta	ttcacaatag	caaagacttg	gaaccaaccc	aaatgtccaa	caatgataga	6000
ctgggttaag	aaaatgtggc	acatatacac	catggaatac	tatgcagcca	taaaaaatga	6060
tgagttcatg	tcctttgtag	ggacatggat	gaagctggaa	accatcattc	tcagcaaaact	6120
atagcaagga	cagaaaacca	aacaccacat	gttctcactc	ataggtggga	attgaataat	6180
gggaacacct	ggacacagga	aggggaacat	cacacaccag	ggcctgttgt	gggtggggg	6240
gaggggggag	ggatagcatt	aggagatata	cctaagtcta	aatgacgagt	taatgggtgc	6300
agcacaccag	catggcacat	gtatacatat	gttaactaac	tgtacgttgt	gccatgttac	6360
cctagaactt	aaagtataat	atataatatta	tatatatata	aagaaattcc	aagaaaatca	6420
aaaggctgca	tacaaatgaa	atactatatg	accattatct	ccttcttggc	tttttcaaga	6480
acaacattta	cttagctacc	acactgaaat	tactattcat	taattttaag	taaaactggca	6540
ttgaaaatat	ggtaaatctg	gtttttcttt	ttttttttga	gacagggtct	cgctctgtta	6600
cccaggctgg	agtgcagtgc	tgtgatotta	gtcactgca	acctctgoc	cccaggttcc	6660
aacaatcctc	ccacctcagc	ttccccagct	agctgggatt	acagctgtgt	gccaccatga	6720
acagctaatt	ttatatTTTT	agtagagatg	gggttcgccc	tgttgccag	gctggcctcg	6780
aactcctgac	ttcaagtgat	ctgcctgcct	cagcctcccc	aagtgttggg	attacagggtg	6840
tgagccaccg	tgcccagccc	agaaatgtgt	ttagtagttt	gaaggggcaa	gtaccagggtg	6900
ctaagagttc	acaggggagg	gatactggac	ttagaccctt	ggaatcaaga	aaagctttct	6960
atagaaagta	ctgcctaagc	tgaagactga	acactgacca	ggctgaatga	aagatgtagt	7020
gctggagagg	atttcagtca	gaaaaaacat	ttgagaaggt	ccagaagaaa	aataaaacaa	7080
tttatttcaa	ggtactgaaa	taaattgaga	ttagggggag	tgaagagtca	aggtctggag	7140
aggtattgag	ttgtcaagaa	atacagtttg	agaggtaaa	gccatattag	gaagtaatat	7200
atttttaaac	ttttgaagct	tattctaagg	ctaataacaa	agtgtgttag	gcaggaaagt	7260
tgcagttatt	ataaaacttt	cagaaagatt	tatttaaatt	aattgggaatt	gagaaagaaa	7320
ggcattattt	tacgtataaa	aaatatttat	ttgttgtgta	aaaaataaat	agtttggcct	7380
agaggagaac	agggccagta	aggaagagga	gaaatgcata	gatataaagg	gtgttaggat	7440
aaagcataag	taagacttaa	tgataattta	aacttaagga	acaaggaaaa	tgaaggagta	7500
atgaatactt	gagtcaactt	gtagatagag	gcaccaatca	ctcaactcag	acagagtaga	7560
ctagagggtg	ttttggggaa	gctaaagact	ccacttttag	atctctggaa	tatccaagat	7620
gggaattcaa	ttaggctatg	atttcaagtt	caaacaagag	aaacgggatg	gagatatttt	7680
tagataccat	ctgataatga	ttagtgtatt	aagccatggg	aatggatatt	ttgtcaatgt	7740
tcttttgaga	gctgtatcct	tcogtagtta	tctgatgggt	tttttaaatt	gcgtaaaata	7800
atcagtaatt	tagaagtttc	tataaagaaa	gttcggctac	aaatacttat	cctttctctt	7860
tcattctcaa	accatgcatt	aaccacctat	taattgtgtt	agaaaaacag	taagtaattc	7920
agggttcatg	ctgacagaga	agaccacat	aagcaaacct	cgctataata	caaggagcta	7980
aattccccaa	tataagaatg	cagtgtgtct	tgtgggaata	tgagaaaca	aagataagga	8040
aaacctaggc	agaaatagtt	caattgattg	ttatttcata	ctgagacaaa	tttgccaaagt	8100
ataaactata	tcctattaca	aattttagat	ctttgggcca	tgggtcagct	ccacggcgac	8160
acctacactg	aagttgtatt	taagctatct	gtctggatca	cttgggtcac	aaaaatgttt	8220
ggattacaaa	aaggaaatat	catacacctt	ggaagaactg	gctcattagt	caagatgctg	8280
cttaccagtt	gcctgggttg	gctgctgttg	taaagsaata	tcggttgga	agccggggct	8340
ggaggtcggg	gagaggctgt	cttctatata	ttacatttgc	tatgtctttg	gaggaaagat	8400
gctttcggga	gtgttcctta	aatttaaata	tcttttctct	ttggactcaa	gggtatgttt	8460
attcgacttc	atcatatcta	atctcctaga	atattattaa	cctgaagtta	ttgtcagtat	8520
ogtatgatat	ttattgaata	acaatatgct	ctgatactgc	atatgaaatg	aaataacaaa	8580
tgaacctccc	ttactttaac	ttaactacta	attaattaga	aaagaacctt	cactcaaggg	8640
gtttggaaag	gtccaataag	ggaatttaaca	gatgagaatt	taagagtacc	tgaacccatc	8700
tcgatttgag	atttacgtga	tgctgggtga	ttagtataaa	gaacagaatg	aagtcagctc	8760
aggaggcatc	gtgactctga	aaggaggag	atatttcagt	gaataattga	gaggattcag	8820
acaaagtcac	atgagagaag	cagagacaga	cagagggcag	gagtcagtgg	atatgagaat	8880
gatactagga	agatgaagaa	tggtactgatt	tgtaaaagaa	ttgagaccca	agttatactg	8940
tgtaagagaa	taagttaaaa	ccattaaaaa	atattgggag	atcctgaatt	gctttcaagc	9000
ctagtaattt	gatgtaagaa	atagtgaagaa	atgctcttga	tggtgtgtct	gtttcaggga	9060
aagttttgtg	gagacaatga	aataaataaa	atagcatatt	tagatcaaga	ggctgggttt	9120
ctttttcaca	ggaaacatct	aagaaactgt	taaaaggagc	aaggtttgag	tgtagaagac	9180

tgaaataggg	ctatggtaat	aaattgttac	cattccgaag	attatcaaat	tatctcccag	9240
ccaaaaaaag	agtttttaaa	caaacatata	aagatatata	atataataaa	taaactagta	9300
tgttttattc	agattcaacc	attttaacca	caaagaacaa	aacaacctca	actgagagca	9360
ggcttgattt	atgtaagcaa	agatttggat	ttgaggcagt	gaatatagga	ggcattggaa	9420
aagtttggga	aaaggctggt	gggctttgtg	agacaaaagt	accatgccat	gagaattgtg	9480
aaaataatgt	agatgtggaa	tcctgtaaca	atggcaatta	gtgatgaagt	ggtgacatgg	9540
gaataaaag	aaatgggtca	acgtggatca	tttttagaag	gaaaagcact	gaaacttttg	9600
aatgagatta	tgaaaaaaag	aggatgagaa	attatgtttc	caaggtttat	tgcaaaagct	9660
gtgatgatgg	cataacctcag	atagaaatga	attattttta	aggaaaccag	tttttagaga	9720
aagataatga	atccaatttt	gaacacattg	aattcaaagt	gacaatccgt	catccatgca	9780
gactaccatg	tagacagata	accaaataga	atagaaacac	ccaccatata	taataaaacc	9840
acaggtgcac	tcattgagtg	aaccagtaca	gataaactat	attgtagggg	ctctcacatc	9900
tggaatagg	aagactcaat	gaaggcaata	aggggaagtt	aaaaagaata	agataaaactg	9960
gggaattcaa	gaagatcttc	tttggcttcc	ataagatcct	agactgtttc	taggaagcag	10020
tcacagttca	gccatggccc	agaatggtga	tgaccgaggg	cagggccatg	gtttggtgct	10080
cagagaccat	ttgccctcta	ggagcaactg	atggagagaa	taaattgggg	agagaataaa	10140
tctagggcag	ttgaaggaga	gagtaaattg	taaaccaaga	atgccatagc	taaaagggaa	10200
gcagcaggca	agtttgcact	tttgaggaga	aattctagga	agggatttct	ccacatttcc	10260
attttaggcc	agaagtcttt	gacatgctag	agagaagatg	ggattttttt	tttaatgaat	10320
aaatgagata	ttgggcattt	ttggcatttt	tgggcaagaa	agatttcatt	ttcaagacca	10380
aaatttaaac	agcaatattt	agcacaaaaa	atgcttgtca	aattcctgct	gaatttgaaa	10440
gcaacagcat	attaagcttt	tttcttcttt	cctgcaatag	ctgatgttca	cgatacagta	10500
ttcccttgcc	ctggtgggag	tatgtcctcg	cttatcctgg	tggaatatg	ttcccttta	10560
tctgtgtggg	atatactccc	tcttatccta	gtgggacatg	ttccaagact	cccagtggat	10620
gcctggaaacc	atgaatggta	ccaagcccta	cacacactat	gttttcatca	tatacatata	10680
catgtatgat	aaagtttaat	ttataaatta	ggcatagtta	gagataaaca	ataatagcta	10740
ataataaaat	agaataatta	taacaatata	ctgtaataaa	agttgtctga	tgtggtctct	10800
ctttctttca	aaatatctta	ttgtactgta	ctcacccttt	tttgatgatg	tgagaagata	10860
ccatgcctat	gtgatgtgat	gtgatgaagt	aaggtgaatg	agaaaggcat	catgacgcag	10920
ggtggcctac	tattgacctta	ataagcaggt	tgtgtctgca	gcattggagac	tctggaaaaag	10980
ggatgattca	tgtcaaggga	agaattgaat	gggacagcac	agattttatc	acattactca	11040
gagcagtgca	caatttaaaa	cttatgaatt	gtttatttat	agaattttcc	atttaattatt	11100
ttgaaaaggc	agtgtgccac	aggtaattga	aaccacagaa	agtgaactg	cagatgaggg	11160
aggactactg	tattcttaga	gcaattgtat	atcattggaa	tattattgat	agtttaatga	11220
aaaaacaaat	tagatcaaca	tgggntggta	ccattaggaa	agatggtaga	ttttcaattc	11280
aaattcacaa	ctttcaaaga	aaaatattgt	gagtctctgc	tatgagatag	atattgagaa	11340
atggtgtaga	tatatattcg	tctgttttca	tgggtgtaac	aaagacacac	ccgagactgg	11400
gcaattttaca	aaagaaagag	gttttaagaga	cgcacagttc	cacgtggctg	gggaggccctc	11460
acaatcatga	tggaaggcaa	gaatgagcaa	atcacgtctt	acacggtttg	cagcggggcaa	11520
agagagagaa	cttgtgcagg	gaaactcctg	tttgttaaaa	ccatcagatc	tgatgagaca	11580
tagtcaactac	cataagaaaa	gcacaggaaa	gatccacccc	atgattcagt	tacctcttac	11640
caggttcctt	ccacaacgtg	tggaattgt	gggagttaca	gttcaagatg	agatttgggt	11700
ggggacacag	ccaaaccata	tcaagatatc	agcagaaatg	agcaaaaaat	gagtgatttt	11760
aaaattactc	ggtttattac	tgacattctg	atgtgcattc	cttaagagat	aattacattt	11820
tacacctaat	gocaaaagttc	ttacagttca	aactctagga	gcaagagctg	ccaagaacc	11880
attgccaaaa	agttacaggg	agagcctcag	tgtttgtttg	gactctccag	ctoctacctt	11940
ccoctctctc	tttcttctct	cttctctctc	catcattctt	cccagccttc	ccaggttctg	12000
gggactcagc	ttccatgatg	ctgcctcaga	ctttctatcc	atacgttatt	attgcactgt	12060
gtctcagagg	agaagaggga	ggccatcttg	agaaggtaca	gggcacctca	gcagtgtctc	12120
ttgctctgtc	attgtccttg	cttctctgat	cctgccaggg	gcacacctgg	agaaacttag	12180
aaaacccttc	agaaaagagt	aacagagaaa	actcaccaaa	tacaagccat	tcaatgtgcc	12240
atcataggtt	caaccacaaat	cttgtttcag	gttacagata	gagaattcca	gtgcttctta	12300
tcagcttttac	gaattttatt	catgcttcag	aaagttgcta	aacagtgtaa	atgaatattt	12360
gggagaagaa	taaaaagaact	actccgaaaa	tctgaaaaga	gttaaaaata	attccagtg	12420
actttgatct	ttcctttata	aatacaaagg	atttatggtc	ttttgcatca	gccaggggtg	12480
ccactgggaa	gtcagagttt	atctatattc	atcttcacat	ttattacaac	cacacacagc	12540
atttatcata	cttgatttca	gacatgatgc	tattattatt	ttccactggc	atggctttta	12600
aaccccaaa	cattagagac	agatagatag	atagatagac	agagatattt	atacaataaa	12660
agaattatat	atataaataa	aatttcttta	tataatgaag	aaattoatat	aaaggaatat	12720
attatatact	atataatggtt	gtataaaaatt	atataattata	cattatatat	tatatgtaat	12780
ttacatgttt	ataattttata	attatattat	atataattat	ataattttata	aaattatata	12840

taaaaattgt	gtatattata	aaatatgtaa	tataataattt	ataatatatt	ataatttata	12900
tataatacat	aatatataatt	ataattttata	tataatacat	aatatataat	atattataat	12960
ttatatataa	tatataatgt	attatatata	tgatgtattg	tatatataca	tacatatattc	13020
caccaatgtg	attttaaaag	actaccataa	actcactgaa	aagaaaacat	aacagttgggt	13080
ttccaagaca	tgaggacatc	agttggaata	acgtttcttc	catagaatta	aatgcattctc	13140
tggtgtcttg	gaaaggcaat	ttcctctcaa	atgggtaaag	tgtttgact	aatactgtat	13200
aaaaatgaat	acaataaaat	aggcgcaaaag	aatctcacia	cttaaaacat	cccatcttta	13260
ttctatagct	gtaagctcag	agagacaatg	atcccatgat	tttatgggag	tagaaaaaca	13320
ctttattgag	tgttttctat	gtatcagaca	ctgtgtacat	tttgctatgt	tgtcccgatt	13380
ttattaaatg	cagtcctttt	cctgttcagt	gtcaccattc	ggaggttttt	ttcaagacac	13440
tctggtgctc	catgagatct	aatctccatt	aatgtgacat	atatgctgat	attatgtaag	13500
tgtgtgtgca	tacgtacact	gacacacatc	gtttgcattt	gtaatttttt	tttaccagat	13560
tgctcataa	aagattaaaa	tggaagatta	gactccattc	agagtttgac	aatgcctcaa	13620
taataaatca	gttctcatca	tttgaagaac	tatgaggtgg	aacaaaaact	gatttgagat	13680
tcataccgtt	gtatacctac	tctttaaaaa	aaagatgcaa	tttacacatc	tgcaaatctt	13740
aggaaatggc	ttagagaatt	tgcatgattt	gggcacatga	agccatttta	gcacataaat	13800
tccttccttt	catctctaac	gggtgtagaa	aggatacaaa	catatttttt	gcagatgatg	13860
tgaattattg	actggcataa	atagcaactc	acactaaatg	cttcagaaat	gtgtttttta	13920
attgagtga	gatggaataa	aaaaatgcag	gaaaagtcct	gttctatgag	agctgaagat	13980
atggaattag	atacagtatt	tgtgagtgtg	tgtgtatgtg	tgtgccatag	ctaattattat	14040
aatttttgggt	tggtcttcct	tgtccattta	atgtttgcag	catttggtga	tggattttta	14100
ctaaatgatg	acaaaataat	gaagcagatt	agtaagaaaa	cctacaaatt	gtttactaga	14160
tcttcactat	ctatattaaa	tagatgaact	aaaaggtaaa	tatggatttt	tacctaataa	14220
agaaaaactac	tagaattttat	tgcactttta	aaatatcatg	tagtgtgttg	taaaatgaac	14280
ccaatcagat	gttctcttaa	aggactatat	tagggatttg	taaatagttt	tcctaaaagt	14340
ggattttttt	tttttgcacc	acaattcaaa	taatgccaga	attgccagct	ataaaccaaa	14400
tgtgaagtcc	atgagcttga	aaagcccata	gtgttcacgg	ggatcagtc	ctttttttac	14460
actggaacaa	gccccagaaa	gttccagcca	gtaatgaaag	atggtgcaca	atggtcttga	14520
cagaaccgtg	atcaaatctg	gtcattgtga	gtcttctggc	cagcgtcgat	tttctaccac	14580
atgtgttttg	cggcctccat	aagatgttaa	aggaacactg	aatcattggc	aatccataat	14640
actcagattc	ttggcttaga	gagattattt	accccggggt	gcctttctaa	cataaacact	14700
agccttgggt	atagtaacag	ggagtacata	gaatagtcca	actcatatga	atgcaagaaa	14760
ttataatgtg	attaattttt	ttggtcattt	caattttattg	aaggtaattc	agttggctct	14820
gtgaatgtat	ataggaaatt	tagaaatatt	aatatgtttt	tatttttttc	ctctagggtta	14880
actgaagtga	atcaaaactat	tggattataa	ttccatgaag	gtaagaaaga	gctccagatg	14940
gttttcatcc	tatcaccagg	acctgcacag	tatgagtacc	tggtgcataa	caaactttat	15000
taattatctt	ttctctcttt	ctttccccaa	ccatgtgtat	acatacacac	gcacacacac	15060
atatatatgt	acatacacat	atatacatgt	gtatggatgc	ataatcacat	acacacgtgt	15120
atatatacac	atatatacat	gtgtatgtac	atactcacgt	acacatatgt	gtatatattc	15180
acatatatac	atgtgtatgg	gcataatac	atatataata	catgtgtatg	ggcacattcc	15240
tctatatata	cacatacaca	cacgtacata	caataaataa	atggttgaat	ggcaaaatga	15300
ctaaatggat	ttagtacaac	ctaattgat	gccacagatt	tgtcaaagag	ccctggaatt	15360
gagtaaaaat	gaatgaacta	aattccaaga	tcttttacaa	cctatgattt	cattccagca	15420
gaaatttgta	aatatgtgta	ccatagattt	atacaaaacg	ttaatgggtg	cattattttg	15480
aatagctcca	aatgggaaaa	agcccaaatg	tcaatcaaca	gtagaatgga	tggatttttg	15540
tatatccata	gaatggaatt	ctagaccaca	gttaaaatga	gtagattatt	gctacaccca	15600
gcaacatgga	tgaaactcaa	aactccttga	gtggataaag	ccaggctgtt	tacgattaca	15660
ttctctgtaa	gaacttttat	tcaaatttca	aatacaagca	aaactattct	acagtgaagc	15720
aagtcatggg	ttaccttttg	cagtttagtgt	acccagcgtg	gggcaggagg	gaagtttcta	15780
ggtgatgaag	atgttttaatt	tcctgatcta	atcctagcct	tactgatggg	tttactttga	15840
aaaattcatc	aagggtgacta	tgcactttct	gtaggtttac	atgaaaattt	caatgacaca	15900
cttgattgta	tgacttttga	tattactttt	cttttctgtg	ccttaatttc	ctcaattgaa	15960
aagtgaggaa	gagaatgctt	acctctaagg	attgacatca	agattagtg	ctgagcacgg	16020
tggttcacac	ctgtaatccc	agcatttttg	gaggcagagg	caggcagatc	acaagatcag	16080
gagtttgaaa	ccagcctggg	caagatgggt	aaaccccgct	tctactaaaa	aaatacaaaa	16140
aaaaaaaaaa	ttagccaggc	atgggtggcat	cgccctgtat	tcccagctac	ttgggaggct	16200
gagtcatgag	aattgcttga	acccgggagg	cagaggttgc	agtgagctga	gatcacgcca	16260
ctacactcca	gcctgagcaa	cagagcaaga	ctccatctaa	aaaaaacaaa	acaaaacaaa	16320
aacaaacaaa	caaacyaaaa	gattagatta	gcaagtagag	tatgtgtaaa	accacctgga	16380
aaaatgcctg	gagtggtgtg	tcagtattta	attgtagaaa	cagttacott	cataaacaac	16440
aggacattgt	gggaacatct	gttaattacc	gaacagctct	tttaggacct	cctgggggtt	16500

gtggatagta	atTTTTatta	ataacacctt	tatatgggct	gaaatcacaa	tgatactgtg	16560
tcaaattaac	cagcagttaa	gagctatgct	tcttcttttt	gccaatTTac	ttgttttatt	16620
tgTTTTaat	tgacaaataa	aattgtgtgt	atttatcttg	cacaacaagg	tgTTTTaaaa	16680
tatgtataca	ttgtggaatg	gctaaattga	gttaattaac	atatgcatta	cctcacacag	16740
ttatttttagg	aactacgttt	ccagaagtct	gtaaggtaga	tggaattgata	tttttaacct	16800
gattcattca	ttcttcattt	aacccaggaa	ttttttgagg	gtcagtttca	gagcatgcgc	16860
tacactaagt	gtcgggatta	ctaacatcaa	ggaacagtta	ctgacacaa	aaatcgatgt	16920
gaagtcctaa	aagagagaac	atagacagac	acataaaaac	atagtataac	acaatatatt	16980
aaaataataca	caacaaaaca	tataagataa	acacagaaat	gtttggagca	atgaattaga	17040
tacttaactg	gtttagaaat	gttttgaaag	aatgtatgct	aatatgtctt	ttgttcaaaa	17100
tttattaatt	atataattaa	cttttaatta	tgataaaata	cataatataa	aattttattat	17160
cttaagtaot	tttaagtatt	cagttcagtc	atgttccacta	tattttatatt	gatgtggaac	17220
agatctctaa	aacatttttta	tcttgcaaaa	ctgagactct	acaccctttg	aacaaccttc	17280
cattaataatt	gtttacatgt	taacacattt	catttatatta	taatgtatta	tcgataacat	17340
atttatatata	aatatatttta	taaatataca	tgaaatttgt	atattttgag	tattagtttc	17400
aaacaatgct	acatatatct	aaggcttcta	gtcctatatt	ggacttacag	tggaaatctg	17460
ggaaattggc	cagcatttca	gtgtgaaggga	ggaaggatgc	cgtgttctct	cccatctgtt	17520
tcagaaaacc	agtgtggaga	agatgcataa	taggcttgac	gcatttttta	taatgatgtt	17580
tatatTTTca	ttttcttttt	catagtcaca	atgaaagtga	aagcaaagga	aagagctaatt	17640
gtccacacagg	ccactgggta	atagcgaact	gtgaagggtg	tggaagtggac	tcactgggag	17700
acgggagccc	atggaccctt	ccccagctcc	tttgctccca	cagcaatctt	ctcccttctc	17760
tttctttgtc	tcacagtcta	gaagagtgc	atttcaaaaa	ttatagcagc	agaggtgttt	17820
ttacctatag	caaaaagtgt	acatgtaatc	acattatgac	atactacaaa	acaggtaaaa	17880
acagtagtgt	ttagtacttt	tataaattca	aacttagcaa	ttttccttat	ttcaacatta	17940
atatcaatag	gcattaatgc	cagtttttta	atgccagtga	tttttaacct	gctgaaattt	18000
gttttataac	ttactgcggg	gtaaatgaga	tatttagtta	tacttttgaa	agtattgtct	18060
gtgtataaaa	caccttgaaa	taaaagtgtg	aagatgtcct	atgagttgct	gtgaaaagag	18120
tagttgtgaa	agtttcaaaa	tgacagatgc	attttaatgg	cagagtagga	gttcatttat	18180
ccaaaaaaga	gggaaataca	gctcaaagag	taaagcagat	gcagagaaca	gcaaacacag	18240
gtcagagcat	tgtgaatctt	cggctttgat	cttgagtacc	tagctagata	ttactgtcaa	18300
cggctaagat	taaaatgtaa	caatatcagg	ggaaatgcac	ttgagaacag	taaaacattt	18360
atTTTgtaca	tatatagctt	gaaatgcata	tgggaaatgt	aggtggagat	tttttagtagg	18420
caatagcata	tgtgagtttg	aaatgtaatg	cggatatata	atctaaaact	ggttctcaat	18480
tctgactgca	cattaggaaa	tttactgctt	aaaatgacta	aatctcaggc	ccaactgcag	18540
attgattaaa	ttaaaattta	tggggatgtg	cgcctgaatt	ttttttttca	ttaaaaaaa	18600
aaaaaaaaga	gtccccaggc	gattataaga	aactgcoagc	tttggaact	gctggcctac	18660
aattagtgtc	aacgtcatca	ttgtgtgggc	tggtgttgga	tggaatgaga	caaggatgga	18720
tgagaacatc	tagagagtgc	aggttggtcc	tcttcttgct	ttctatttgc	ccttcagatc	18780
tgccctttca	tgtagtctgg	gcaccagagc	accctgagcc	cagtggactg	caccagtggg	18840
ctcttccatc	ctcagggttt	tggttaagcac	catctagagg	aggcatagag	agaaaatcag	18900
ggaccacagg	aggtgaaaac	aggttttgtt	ttccagctt	cctttctgct	gggtcaaggg	18960
gtaactgcag	gttgagggtc	atcctcagtc	aaaggccaca	gctgcagcag	gtgcacctct	19020
tcacacagct	gcttctagag	tctgcttagc	tgcttcaggc	caagggatga	tgaggtctcc	19080
caactgttgc	tagctctata	gttctcagcc	atcttcagtt	gattcccccta	aatcctgcct	19140
gccccttctg	agacagtacc	ttcattaaat	attcctcatt	ggttaaaaag	tagtgacctt	19200
gtgctgaata	ttctatgcca	agcctctgcc	tgctgcagta	acataagaag	agttgtggct	19260
gaagatggaa	ccaagagaga	aaatgacctt	taaaagatgt	aacgaggtgc	ccagggaagg	19320
ttttggcaat	ggaatctgag	catgtagagt	ttaaatgtaa	gaacggaaaa	atatctgttg	19380
tatttgaaag	gtcacgaaag	gcttcttccc	gagaggaatt	ttagtggtgc	gctaacaaaa	19440
atcagatggc	aggggctggg	gcatcaatgc	aagtgaagaa	atgaagggtg	gttattttgac	19500
aagtttgatg	ataggaggaa	taaagaattt	ggatttatta	aaaataagac	atggagctaa	19560
tcaattgaga	ttgccaaaact	cttttaatga	aaagttagctg	tgctctatct	ccagcagttg	19620
caacagagca	ctaataatat	caccgagact	attgatgtga	gggcaatata	catactcagt	19680
gctgtaaagg	cttagatttt	tgatagctac	tcaagggaaga	acacaaagaa	ttaagtcttg	19740
atTTTTaaaa	tgaacaaaga	tataaagaga	gaaaatacag	aatgtatacg	aataatcaaa	19800
acatatgaaa	cgacagccat	ctctggtaat	caaaaagaag	ctactacaag	caatgatata	19860
tgTTTTgcct	tataacttca	taaagattca	gaaaatgggc	catagccact	attggcaaga	19920
aggcaataaa	agtccatttt	tggtcagaat	gtcaatttat	gtatcttttc	tgaagtctgg	19980
tctttgcact	cttacccttt	taaagaaaac	attaattgtt	tcactgaagc	tctgagttca	20040
tatggccatg	ctctcctgcc	caggtgttct	ctcccttaag	ccttctcttg	agcacttatg	20100
tgccactcat	catcatgact	tcagtcttta	aagtttcttt	aataccactg	gactataaaa	20160

taggactgac	gaaaagaccc	attcaacatt	ccattgaagg	attaaatgct	atgttttaag	20220
atataatttt	ataacataca	tctatcatga	aatccattga	ctcttcacaa	ttagatgctc	20280
ttacatagaa	aacatagtta	cttacaaagt	catgggtgaa	atgaatctct	gctgtaacgc	20340
cattacatat	ccctaaaaca	tatttaagaa	ttatgaagag	gagcattatt	gaaaacaaaa	20400
tgtattgaca	ctaaagagta	actattttta	aaaggtaata	tcatgatact	catatgaata	20460
tcaaataaat	gcatgtcaaa	tatttttatt	aactctttta	ctaatagaaga	aatgggtatg	20520
atgctacacc	caattcaagg	tagaattcaa	aattcaaaaat	aaacacatat	atacataaat	20580
caggaatttt	gaaatgaatg	tacaaaatga	tgcaaaaactg	gctttttcca	tggagattgg	20640
ggtcgttgtc	caggaatcag	tagttatttc	cagggacaaa	attcatttgc	atttctatgg	20700
gcaagtgtca	tttaaactgc	aatcagttat	ctacaattta	tggaaatcgta	accactttta	20760
aaaagatgaa	aacagttaac	ctgaaagttt	ctcctgcact	aaatgagaaa	gaagctcagt	20820
tattttttaa	tttgcttatt	tttgtcttgt	tcttttttaa	tttcacaata	tttgactgac	20880
agtaactttc	ctgatagtga	taaatcaatt	tgtgccacat	ctcatggcat	atcagctttc	20940
tgtgtctttg	tccatctcaa	acaacccac	atgacctac	caacatgagt	tttatctccc	21000
ttcaaactgg	attaatggta	tccatgttac	atgactttat	acgtgttctt	cactttacag	21060
caaatagtta	tgaacaatta	atggaacaca	cacacacacc	cagataggaa	tgggaccctc	21120
tgagttttgt	tcacagagtc	tatactcttt	tctatttgtt	catataatgg	tgacaattca	21180
attaggtggc	ctatgtccac	aatccatggc	ataaacgccg	gtaaatttac	agattctccg	21240
catctatttt	ggcaagtgtc	tcagtgtat	ctgtagcctg	aagagaaaaag	ctaagttttt	21300
cctattttct	tctctctttt	ccttcccttt	tcttttgaat	ttcctaaaaac	aaaggaagaa	21360
aaaggagggg	aagaatttgg	caatgtgtca	gttaatttca	tcaatgtaga	caaagttaaa	21420
ttagggaggt	gaattaccaa	gactggctgt	ttccattgat	gaaattaaat	caggtaaat	21480
gccagtgtct	cacaaatgct	gtctcaggtg	tgactgaagt	tgtgattgat	tggaaagtac	21540
tcacagagaa	aacaagtaca	ggaaaaacag	cactattcca	aaatgcagaa	caaaaccaca	21600
catatttagg	ctttacagtt	ttcgtgtctc	ttcatattct	tactttgatg	tatttctgcc	21660
ttttaattta	gsccttttct	aaaaaaaata	gtttgcttta	cagtttacaa	agtaagatat	21720
gctttaagag	gtagaaaaat	gaggcaagtg	aaaatgaatt	aaaaaaaaaa	aaagaatcca	21780
ggagcaaaat	cagtccacaa	agtttgtctg	tgacatcatt	gtccttccaa	acaaggagct	21840
tcgagtctca	ctccaaacat	tagcagtcca	gacgggaaaa	tgcaagtacc	aaaatcagag	21900
tcctgtgggt	aaaagcaagc	catttaatat	gaatgaaaga	gtcctgtctat	aaatactaca	21960
gaaaataatt	ttcaaatcta	ctatcctaag	gagaacactg	tgcaagataa	tgaatatcat	22020
ctttagaaac	atcttcttgt	aattatttct	taaaatatcc	ttccagacaa	agcaatgaca	22080
tcattacaaa	ctccaactta	gcaaaacgct	ttttggaaaa	aggaaagtgg	aaggacattg	22140
agtgtgtctc	acatatgtag	ctctctgatg	atttgacttc	atccaagaat	cgtgtggaaa	22200
cagctagaat	gggtgtgaat	ttatatctcc	caaacagttt	ttcataaata	tcatttttatt	22260
tgactaagct	ttcaagtgtg	actcaatgat	agtaaagggg	agaacctaat	tcctgacaat	22320
tgcatgagat	gttggtcttc	agtgttgcca	gttaactgag	aaactatggt	aataatcagc	22380
cgcgttagcc	cttttctgtt	agataatttt	attattcgca	attctctttg	atgggtatta	22440
ggtttagaca	attgaccagc	aaactatttt	tttttgaaaa	tgaccacaaa	catgttaaaa	22500
atggatgaac	cctgagagga	gacaggaaaa	cagcacctaa	aaatagtaga	aggatattat	22560
ttctctgccca	ttatggactt	ataacataat	aaacatttag	ttggttcact	tttttagtat	22620
caacactttt	attaattgct	tcttttttta	caaatttgat	tttattattt	tgatacctac	22680
tgatctctaa	tttcttgaga	aagtctgtct	tactttaata	aaatgacata	cagtgcattt	22740
taaacttaaa	tactgtatat	tcataattta	atctgagatt	attgggacat	aatgatataa	22800
gaattttatat	aaggccatcc	tctgaatcca	tattcaatta	aactaaaaaa	ggatgtccga	22860
tttaggaatc	ccacagaccc	atcagtaatc	aggcaagtac	ttaacagcca	tcctatcatt	22920
ggctctctca	gttttttatt	attctgttag	tattagcaga	tgatttctat	aatgttgatc	22980
aaatcagatt	tacgatctaa	gttgagttaa	atctcagaga	atcaaaactcc	attgtggaat	23040
ggaagttttt	agtttcttat	atatggcaat	attttgtaag	atattttcat	aactaacaaa	23100
gtattctcat	tataaattaa	gtacatttct	cccttaagct	ttttaaaaaa	aatgatttta	23160
gaaaattcct	gaaataagta	tcaaaatgca	agattgtaaa	atacgttttg	ataaatcaga	23220
aacattctta	ttaaaggttt	tctttattga	taagttatac	ttctatgttt	ataaaatatg	23280
ttgtattgtt	aaatgaagcc	aatgaaattg	atcaagggtt	aggtttttaga	atatttctgt	23340
ccagagcttt	gtcctcccaa	attctgtcgt	gttagaattt	tattttttga	ctacaggagg	23400
atagaagcct	aaaactacat	aatccgtttt	taaaattaca	catgtaaaat	23460	
gggtgtgcaca	aataaatact	ctgtcacaca	aacatgcaca	cagagcataa	acacatgtac	23520
acacattaca	tattaatgct	tcttctagga	cagagattaa	taagtagtag	agaaaaaaga	23580
attacagata	caatctcaaa	gcttgtcata	tttctacaag	gtaaatatat	atctcagtg	23640
tagtgctatc	aaaaccatac	catcaggtta	aagaagtggg	taaatattat	cagaatttaa	23700
tgtgaaaaca	ggattatatt	cctactattg	ctgggtgaca	tagtactcaa	cacatttttt	23760
ttcgtgttat	tcagtgttaa	tgttaatctg	ttaatttaaa	ttaaatcaaa	ttaaataaaa	23820



tctgttaaag	aaaattgttc	aaatagcttc	tccctaaagt	tttctttagt	tcaaaagatt	23880
ctgatgctat	atttttaacct	acagctttaca	aacaaatttg	ggtcttttaa	aacttccatt	23940
ccctcaattt	tagtaaaaaa	gaaccctact	tttataactt	acattaaaca	ttggaaataa	24000
tgtataactt	acttattagt	cttaactttg	ggagagaaga	ggtttaagta	cgtattttta	24060
aaaaaacaac	agaaatgagt	actcatgata	aatagtatat	gcacctaaaa	tatactaaaa	24120
ttaatgagta	caaatactca	ttcagctctc	aaagtaactc	tcccaactga	ataaacacgt	24180
ttaaaaatat	tttaatgtat	ttattttgtga	gttacacaaa	tccattagtt	taatatttag	24240
ataatatgac	aatgatatat	gagacacaa	ttattttaag	aatatattata	taataaataa	24300
ataaataaat	aaataaaaaat	aacataacct	aagtatacct	tagaaaaact	atggggaaat	24360
gttaaagata	aggcttggga	actttgccag	cacagtgtgg	gtatcatagt	gtaggtattc	24420
ttattttcaat	taaactagtt	gctttcagtt	taaaaagtag	atgattttac	atacaagttt	24480
gcactttctag	cttctctttt	caaaatgagt	tcaggccttc	ccatcctgac	accttgaaaa	24540
agctgtggcg	aggcggggcg	cgggtggctca	cgtctgtaat	cccagcactt	tgggaggctg	24600
aggcggggcg	ctcacgaggt	caggagatcg	agatcatcct	ggctaacatg	gtgaaacccc	24660
gtctctacta	aaaatacgaa	aaattagccg	ggcggcgggtg	cgggcacctg	tgggtcccagc	24720
tactcggag	gctgaggcag	gagaatggcg	tgaaccgggg	aggcggagct	tgcagtgagc	24780
ccagatggcg	ccactgccct	ccagcctggg	gaacagagag	agactccgtc	tcaaaaaaaa	24840
ccaaaaaaac	aaaaaaaaaa	acaaaagctg	cagcggaggc	gtttctgaag	atggcgggta	24900
ttgggcactc	tttacttcct	tgactctact	cctgcactct	gagactgacc	gtataacatt	24960
tccgattttc	tttgatttgt	gctattgttt	ttcttatagt	taagggaat	gtgaaatatt	25020
attgtacctt	tgtctgtcaa	aatttagaaa	aatgaaaaat	agaccaaagg	tgggagtata	25080
attcattgca	aaaacttgat	atcctaagct	agtgtgtcaa	aagtgtacaa	tttcgagcat	25140
tattatgaaa	taaatttagt	aagatctatc	accaatatat	aagaaatagg	atacaaagtt	25200
ttatgaacta	aaaagaaaat	tctcattgaa	ttgtttttta	atcaaattaa	atgaagctac	25260
atcgaaatat	ttaaaaacat	gttggtcaatg	atgagtctat	aaagcaattt	agtggtgata	25320
acccctccaa	tttgctgata	tgggcctctg	ctgggttttg	gggccttgaa	tctgcaatcc	25380
ctggtggagg	acaacacccc	tgtttctccg	gacctgagag	ttttctttt	aatctaattt	25440
ccttcacatt	ttaatcatta	gcttatcttc	aaactttttt	aagaatgatc	ttacaaaatt	25500
aaaagtaact	tacaatgtat	tgcattttact	atgtttgaat	aaaacattat	tttgccacta	25560
tcccaaggga	tgtttatat	tttctatttt	agtaagttta	ttgaagacaa	acaggccttg	25620
ggcgcgtgcc	tctgtctctc	ctgccttccc	cactgaatct	ccagcccaca	ttcacaaaaca	25680
aagacaaacc	tgtgtaaact	gaagcaccac	aaaactactt	ttatgaaagc	agaatggaag	25740
gtaagtgtca	taatctgttc	ccccttgtag	tgtagcataa	actttatcga	actcactgca	25800
gcaaataacct	ttgtctttca	ccaagaaatg	gaaattgttt	ggtgtaattg	tgccttacct	25860
tcccagagta	tatttctgga	tttgattttac	tgaattcaaa	gttgaatgat	tcaccaaaact	25920
tottggtttt	caatgcttaa	agtcccaaat	aacttttctt	aaaaatcagag	ctgccaaagt	25980
taggaatgat	tttaatatca	attcccagaga	ggtgtagaag	tagctactgt	gctatttctc	26040
acactttatc	aoccaaagtc	accttcacaa	gcatoctcag	gatgaagact	tcaataaactc	26100
ttaggctgtc	catttctcta	atgcacattt	gtggttgtca	gaaaattttt	tttatattga	26160
tttgaatcaa	ttgatccatg	ttttgcccac	cagaacaaca	gaaaccaa	gtattctttt	26220
tgtagtactt	ttcattttta	aaaacaatta	ctagaatccc	ctgagtcttc	ttttgtcttg	26280
tggctcctcc	cccaggacag	tgtgtccagg	tggactgccc	ttccaaccct	ctcctctctt	26340
tcagacaact	ttgtccatgt	ccctttgcag	aggccctggg	cagagattcc	cgccgggagg	26400
totgtctaga	ccagtgcacg	ctgaagctaa	tgagtcattt	tatatggact	tgggtgatgat	26460
ctcagtgtca	taaaaaggag	attccattat	ttctctttca	aatacagccc	tagctgtctga	26520
ggcagatctc	tctgattctg	tatttttaaat	tgatttttga	atttaaattg	agagatttca	26580
ggtctttatac	tttttagagcc	aaaagagaac	gttcacatca	actggagatt	aggggttaatg	26640
ttccagatca	acggccaaat	agttacgtgg	gcttgaacaa	ctgtcccagt	tggcctcctg	26700
ggacagctgt	ttgctgtaat	atgggaatag	gaatgtcatc	tcccctgcc	ttacaggact	26760
cttgtgagga	tctactaaag	aatacccata	aaagtctttt	aaaaagggtta	gagaacagaa	26820
gatgtcattg	gtttatacgt	gtatcctgat	tccactcctt	ccatatcatt	gtctctctgt	26880
gaaaaccaga	ctttcttaaa	agctcttgta	ttattatttt	ttttttttcc	aaacaccaac	26940
caatactctg	acaccatctg	tgtggacagc	agttcaattg	gattctaact	aaccagagtt	27000
agagaaaacc	ccacgggatg	agggcccaat	gcacaagat	tgcccccata	tcacatgcca	27060
gccacaaaaa	gagtgccagc	cttcctggtc	tccctgccag	cccactacaa	acacagggtt	27120
cccacccttc	ccaccctcc	aaatgtgaca	actttttaga	acaactcata	gaactcagga	27180
gaacactaca	cttactgatt	ttttnnnggtt	tnnattatga	aggacagttc	aggaccagtn	27240
caaatggaag	tgggtgcacag	gacgggctgg	gtatggggta	ggggaatgac	tgggagcgtc	27300
ctcaccctct	ctggggggcac	caccctccca	gcacatccat	gtgtttngcc	agcttgggaag	27360
ctccttgaat	ctcttttngc	tcacgagttt	ttataaccca	atctctaata	ccctcctttc	27420
cctggagggtt	aggggggtata	gctgaagatt	tccaacgcct	aatcaccatt	tgggtctttgt	27480

tgtgaccagc	cctatcctga	agctatacag	agggccgcac	cctaggttac	ctcagtgcca	27540
taaacacagg	tgtagtagaa	aggggcttat	tatgaataac	agaaggcact	cccatttctg	27600
ctagtcccga	gggttttgg	agttctgtgc	caagaacagg	ggaaaaagat	gaagtacaat	27660
ttttattaca	ccactgctgt	ccatactgg	taccagcatt	ttctttgggc	tctcctcaaa	27720
tgacaccact	ctccgctgcc	ctgtcacttc	cctgaacagt	tctcttggtg	atcttctctc	27780
ctcgtcctac	tttccctttc	atgagcacaa	gaagtagtca	agcgggtgaac	cctttatcgt	27840
tcttgaagca	cctctcactt	tggagacttt	aataccacaa	cctgcttggt	ttccttccat	27900
gcttctgact	cttctgctt	tgtctccttg	gagatatatt	agtgaagaaa	gtaaaatgtt	27960
gcaactgggt	taaatgataa	tttgatatac	agatatatgc	attctccaag	gcacagaatg	28020
tatttatatt	gctaagaaat	gtgctaaaaa	gtggagacaa	ggcctgaaga	atccctgagc	28080
agacaacact	agtcaggcct	cataagtga	ctcaaccttg	tttgatttgc	aaacataagt	28140
gaatgtcagc	ttgagctatt	tcttgtaaat	gcctgtattc	aaaacagaa	ttaagcttaa	28200
ccagtcagaa	gcagtcctact	cacttataat	cctataactg	gggggacttt	ccgaaggat	28260
agatcaaaata	aggcaactgt	ataactgaaa	ccaaccaa	gctttctctg	atttacttct	28320
gtgtgttct	ataaaagcct	cctctttctg	tttctcagt	aaagctcctg	aaccagttct	28380
actttggagg	tgcccaattc	atgaattgct	tgtctcaata	aactattctt	ttaaaattta	28440
ttgggcctcc	atttactttc	agataacagt	aattatttgc	atgcctatgg	aaggaaatg	28500
tggtattact	gtcttttttt	tttttttttt	tacaatttaa	gttctaactc	tacagaattg	28560
taaaataaca	taagaattaa	cccaaaatgt	gcattcctat	caagtcagat	aatcaccag	28620
aaatcctgct	aaaataattt	ccagaatttc	actgaaagga	cacctagtca	tttcttttgc	28680
ccatttctctg	acagtatctc	cactagtaat	aaaaataata	tgaaagaaa	attcttaatc	28740
tcaaacaggc	caaggatgcc	cataagggtt	gacctgatat	ttacatagtt	gcagaagaat	28800
ttaaactaac	catatgggca	tctttgagtt	tattctgtag	ttcttaatac	atgaattatt	28860
aagcaactac	taagaccaga	gaataaactt	atgactagga	gttttcaaaa	gaaatctttg	28920
attggtcttc	taaaaatccc	tatgattttc	ttttctgtat	tagttcattt	tcacactgct	28980
ataaagaact	acctgagact	gagtaattta	tgaagaatag	agggttaatt	gactcacagt	29040
tctgacggct	taacaggaag	catgacttag	gggtctcagg	aaacttacaa	tcatgccaga	29100
aggctaaggg	gaagcaagca	catcttacca	tggcagagca	ggagacagag	cgagtgaaca	29160
aagggggaag	tgccacacat	tttcaaacca	ccagatctct	tgagaactca	ctcactatca	29220
caagaataac	aagggagaag	tctgcctcca	tgattcaatc	acctcccacc	aggccccttt	29280
cccgacacat	ggggattaca	gttcgagatg	agatttgggt	ggggacacag	agccaaacca	29340
taacatatcc	taagacgagg	aaactgacca	caagaaactt	tctccagcag	atttcacctg	29400
catttccctac	attcaactga	atcctttaaa	ttgactgagg	cacatcctgt	accaaggaga	29460
gacctttggt	actgcctttg	atcctgagac	cttctgagcc	aggtaacttg	gtgcttttcc	29520
tggggaggct	ttgaggcggt	tactccacac	ataatgtcaa	acgtgtttta	taatagtggg	29580
atagtcagac	ttgattaaat	gagaattatt	ttcaaatatg	atactaagta	aggctctgag	29640
tgtagaatca	aattgtccaa	ttatatccca	ataaaaagga	agacagattc	ttagtggacc	29700
ttttcaacta	tctatacagc	agtttcataa	aacattagga	aatttagtaa	accttattac	29760
ttctggggag	tcagtgagaa	tcagatacat	ttcttaaatg	tttcatttca	tcaagagcag	29820
agtctatgaa	attaagggtg	agatttgctt	caaatcttaa	gaagaaagca	ctttgatata	29880
tatcttttaa	aactaaaaac	aacattgatg	gtattctaca	atacttgcac	tttctctcca	29940
cttatgtagt	tcattcctgt	tctcaggaac	cctgctttta	agaggaaggt	cttggccagg	30000
ctcagtggtc	cacacctgta	atcccagcac	tttggggagg	cgaggcgggc	agatcacctg	30060
aggtcaggag	ttcaagacca	gcctgaccaa	tatggtgaaa	ccccatctct	actaaaaata	30120
caaaaattag	ccaggcatgg	tggcagggtg	ctgtagtccc	agctactcgg	gaggctgaga	30180
caggagaatt	gcttgaacct	gggaggcgga	gggtgcagtg	agccaagatc	acaccactgc	30240
actccagcct	gggcaacaga	gtgagactcc	atctcaaat	aagtaagtaa	ataaataaat	30300
aaataaataa	ataaataaat	acagtgtcct	ggaaagcctg	actcggtaga	tggttacctt	30360
ctgagagtgg	gttaagtaat	atcaccttag	caggtgatag	tacataagag	accattctgg	30420
aagtataaaa	agttaaagtt	gctggtacaa	acatttccac	aaggctctaa	gagccctctt	30480
ttactgaaga	caaactgtgg	tgtgaatctt	acatcagtat	ccaacatcag	agtaacgaga	30540
aactacctgt	agatgacaaa	gactgaatgg	ctgtggccaa	tttattataa	taaccagtga	30600
tgtcagaatg	gagaagaggg	gaattctcta	tgggtgttag	taggcaacaa	ttccataaga	30660
ttttggcaat	gtttctcctg	tgtggaat	tactatgtac	agccagagta	ctggcaaatc	30720
gtctggacct	tcccatgaat	tatataattt	ctaaaatatt	tatattaata	atttataata	30780
ataaatgaaa	ttgtgtcctt	gaagttcata	catcaaattc	ctaaccacga	ataggactct	30840
atttgagat	aagggttttg	tggaggtaat	gaagtttgaa	tgagatcgta	agggcagggt	30900
actgtgatca	taagaacccc	aatatgactg	gtgtccttat	aagaagaggg	agagacacaa	30960
ggagcatgct	tacaagaaaa	ggcaatgtga	ggacatagtg	agaaggtggc	catctgcaaa	31020
tcaaggagag	agaccttaca	agaaacctac	ctgtccagca	ccgtggtctt	ggatttccag	31080
ccoctagaac	tgtgagaaaa	taggtttatg	ttgttcaagc	caccagacaa	gtgttatttt	31140

attatggcag	ccaaagagac	taatacaata	atatttttct	gtttaaattt	tatctagcaa	31200
tagctgagga	tctttcccaa	tttaataatg	cttgctatgc	agctctcacc	atagtgtcga	31260
ggtaatttaa	gaaatatagt	gcacatcaca	aaacataatt	attttagtagca	tcactctttt	31320
cttaagggtga	aagaataaat	ttttatgaaa	tctcaaaaaga	gatgtagaca	tttgattcat	31380
ttttcttaaat	ttaggataca	atcttagaaa	gccaaaatct	aaataattgt	tggtggagat	31440
ttgggcactt	atttaggata	agatcatgga	tacctaaaaa	acagtgggtg	gttaattcat	31500
caaaatgaca	tcaaaatatt	ttaaagttaa	cttagaagtt	tccacaatta	taaattattct	31560
tagctctttc	caaagctaga	aaatgtttta	tttttagtaa	tcaaagatat	atgaaagtaa	31620
acacagaaaa	attattcaag	tgggatacag	aatctctatc	acctattatc	agaattttga	31680
ttacacaaaa	gattaaaaaa	tgacttttca	cattgttaggc	aagtcacgca	tcattaaacc	31740
aagggaacaa	gcacatcaaa	aacagagtga	aatttgccaa	aatataaaat	atttttagttt	31800
tgtttttagat	taatatgtta	ttacaaaaga	caaaattcaa	tttctaagtt	ttgtttttca	31860
ctatactott	tcatattctg	gaaagatctg	tcacttttct	aaaatacaaa	actaatttctt	31920
ttctctgaaa	gacaaaaata	tacacacttg	tgtttccctg	actctattat	ccttagtaga	31980
gtctcaatca	cccatttgat	taagcttatt	tcatagaaaa	aactgggtta	atgggggaat	32040
aattaagaac	ggatgtcata	cagtatttac	tgtattacca	aagggtcacag	gaaacacacc	32100
tggtggccgt	gtcatccatc	ctttacatct	gcttgccatg	gcaaaaaata	atatactcat	32160
tgatatagat	gtatcccaag	agatagtctt	ctggacagca	tataaaaaata	acaagcacaa	32220
gtatgtaaac	ttaaatttat	atttagtagt	caaggactca	actgttatac	attcattaga	32280
agttactttg	ttttctaaga	attctccaat	tttaagggtta	tttaacatta	gtcaagggtct	32340
taagttaacct	aagggaagtta	gaaactaagt	taaagctgtc	acactgcaac	acatagttac	32400
tattgacaga	aaaggtttat	cataataata	atttaattta	atgcaatttt	tcatatttta	32460
cacattgtgc	agaaataatg	ttagttttta	tgatcgatag	accactataa	attcagaaat	32520
tgcagtttga	ctcttatcat	catgagaaaa	gcaacccaaa	tagagtgaat	tgtaggtcaa	32580
attatatttta	gtgtggcttt	atcagaaaaa	acatatagac	tttttaataa	atgaaattat	32640
ttataataatt	tgccctaataa	tggaattaga	taaacttgga	atttaaaaaat	gcttctgaac	32700
taaaggtttt	ataagaacat	ttttaatgtt	ttccaacttt	attggtttat	aattaacata	32760
aaaaatgtat	atggtgggtt	gatgcaaaac	agcaattact	tttacacaaa	ccaaatattt	32820
aatgtacaac	ttgatgttgt	gacatacata	tacactgtga	cgtgacaatc	acaaacaact	32880
tagtgtatgt	atctattaca	taccatgggt	atctttgtgt	gtttgtgggtg	aaaactctta	32940
atatctatoc	acttagcaaa	ttttaagtat	acaatacact	attaagatg	atatttttaa	33000
tctagtgtat	ttaaagtatt	agtttatttt	tctttttttt	gaatttgga	aaattggatt	33060
tatatgtaca	tttatcttta	aaagccaatt	agaattgtac	ctctttaatt	aaagagatag	33120
gaacatttat	atccactaac	tgaggtcaag	acttttcaca	attacagagt	cacaaacgtg	33180
ttagagaact	cacagcctga	tatcttccac	ttcagccacg	agtcaagaat	acacacaaag	33240
atgcaaaaaat	ccaatggctg	agatctcaaa	gagaagttca	gtttcccaat	gtgcataaaa	33300
ttctcaattg	atttgagctc	acaagtaaaa	caaaatagac	aataagaata	ataagccgat	33360
tacgtcatct	tttattcaac	agacagtga	cctacagaga	tcatcgatt	ggcagatcat	33420
aaagccaaat	tctaattgc	ttctctcagt	ggaaaataag	ccatcagcta	tgtcttactc	33480
agcaacaaaa	caaaatatat	agttagtagg	aaacaaagtt	aaagaaacag	agtgaacaag	33540
gttaaatatc	tattgccatt	tgataccctc	tctgggaggc	aggaaaattg	ttcatgggct	33600
taagcaattc	aggaggtcca	tttcttcagc	cccacagttt	acctgactct	aagttgatga	33660
actagacatt	ttggcagata	atctacctct	aaacccatta	acacataaca	ttcccacagg	33720
taaaatcaca	attctcatgg	aaatatcaaa	taaacatatg	tcaaattgta	tttaactcat	33780
cagtttaacg	aaagaaccag	taagatgttg	aaatcagttc	tcagaaattt	gttgggtttt	33840
ttttcaacca	agaactcccc	actgacacca	acctatcaa	aaattgccat	ttgacatttc	33900
atctggatgt	tttaaaggcc	tctcaattca	ctgtgtccaa	aaaggaattc	atgactttct	33960
tttccaaatc	tctttccttc	tcttgggtg	ttcactactt	ctggcaatga	tcccactttg	34020
tgttgaaatta	tttacaccag	aaatctaaga	ctaataaatg	atactttatc	tcactacact	34080
ttcacctgcc	attacatgtt	tattaccaaa	ctctgttgat	tttactttct	aaatatctct	34140
tacctgtatc	tgcttctaac	ttaacaccgt	tcatccatac	catgacacca	ccatcaccat	34200
ttcatctgga	ccatatcact	tagactaaac	cattagccat	atgtttttct	cattgggttc	34260
actacctcaa	gcattctctt	atccctccag	ttttgtccat	atatcagcca	gagtgataga	34320
tatgataatg	tattctacta	ttagaaatcc	gtaatgattt	cctattgctc	ttgggataaa	34380
aaacaaaaa	ctatgaagcc	ccattgatct	taggataggg	acttaaaact	tcttattctc	34440
caccattttt	gtgatagaga	ctcaaacact	tgatctctcc	ctcctcttat	aataaagacc	34500
caaaccttat	aacacttcat	ctctcttagg	ataaagacca	aaatctatga	tacctgtctg	34560
ctctcaggat	agagacccaa	gccctctatc	tgaggtacaa	ggctctctta	caatgctgct	34620
agttacctta	acttctctct	ttcaaatattg	tcatgaacca	ggattccccct	cactttctgg	34680
ctcagccttc	cttaatttct	tcataggttt	ctcaacttct	catcactgca	tcttagaagt	34740
gtataattta	tttattttatt	tatttttact	ttcaatattt	ttctcttaaa	tctcctactt	34800

tgtgcctg	ttcaaaccta	tgattatcca	agtttatctc	atttttattt	tgtcagggcc	34860
ttttccttga	ctgttcaggc	aagaccaagt	tataccctc	tcctgatagc	acttaaaaaa	34920
cattgtaate	ttggccaggt	gtggtggctc	acgcctgtaa	tcccagcact	ttgggagacc	34980
gaggcaggtg	tatcacgagg	tcaggagttc	cagaccagcc	tggccaatgt	ggtgaaaccc	35040
cgcttttact	aaaactatga	aaatttagcca	ggcagcgttg	ctcatgcctg	taatcctagc	35100
tactcgggtg	gctgagacag	aagaattgct	tgaatctggg	aggcagaggt	tgcagttagt	35160
tgagctgaga	tcacatcatt	gcactccagc	ctgggcgaca	gagcgtgact	ccgtctcaaa	35220
aaaaaaaaaa	aaaaaattgt	aatcttacat	tgaattttgt	ggttcagttt	ttttctctac	35280
cactgcacag	ttaattttcaa	gagggcaggg	ccatgtctaa	ttctacttat	tatattccta	35340
ctgaatatta	tgtagttgac	atatattaaa	taattttaat	ttaaataaaa	tgtattttatt	35400
gagtgaatgt	aggacttgct	gtcatataac	atattttctaa	gtctgattct	atcttctact	35460
atactagata	attagttcca	aattttatgt	atatacgtta	taatgtgatt	agactcccat	35520
aaaaaagtat	taataataaa	agagagttaa	tgtgtcatat	cacaagaagg	ccagaagtag	35580
gaaaagtgtg	gtagtttagc	ctctgaatca	gatgggggta	taaattaact	tcttttactc	35640
accagctgtg	tgagcacagt	tatatattta	cagatcactg	aaacattttt	ttgaaattca	35700
gtctgctcaa	gaaaaaaaaa	aggaatgaga	atattacctt	tcacacaggg	attctttcaa	35760
aattactgaa	ttaatgcatt	caaagaattc	tttcagtact	gttcaagcac	ttcaatatta	35820
ccaatgacag	tttcttgcag	tatcttggat	attattcaag	tggacctata	gagtgttagc	35880
catttaaaag	tcacagctg	cttttgctgt	ctccatgttt	attgcagcat	ttaggactag	35940
gttcaactac	tcataacaga	aaatgttact	gtagtggctt	aaacacaaat	aatttatctt	36000
gtgcaaaaca	ttcaggggtg	agcatttggg	gagaataaca	ccaagaagcc	agaccttttg	36060
taacttctctg	cagtgggttt	agcaccttca	ggcgggaaag	tcaaagaaga	gaggagtgc	36120
attgggagac	tttgatgtac	gtatcaactg	ccagaagcaa	tcattttctaa	cagtgaaaaa	36180
ggaggggaaa	tgaaggtttt	tgctccaca	atctctatag	tagaagcaga	caaggcaata	36240
ggggaaggga	aaaaagtttg	gtccatcata	tatcctaggt	gtagggtattt	gaggggtgatg	36300
accataccag	ttaattaggg	ggagatgagt	agatcctcat	ctcttccaag	tataggaagc	36360
taactttgcc	gtaattgtta	ctaatacacc	cttcaaaaatc	accattttga	aaggttttgc	36420
gaaatgagca	tatttttctt	ttaatacaag	gtcagacctt	cagaatgatg	aataaaactg	36480
tttgaggcca	agtaccaatt	cagagctgct	gtaaagtagg	cattttcaag	aagaaggaat	36540
acttctcatg	tttgaactgt	gtaagcaagt	ctttattttat	tttttgaatc	ctgaaacagc	36600
aattagaaaa	gagagaggaa	gatatttctt	tttctgaagg	ctgccaagaa	taaatcctat	36660
tctcatttgt	taaaataaaa	agaggttaata	agctaaagta	agaatacact	aacctttatc	36720
ttgaacccag	cgcctctag	aaaaaaacac	tataaaagtg	aagctaagtt	cagtaatttt	36780
tgcagtcatt	gcctttttgt	tagctctagt	tatctttcaa	gttttattga	taataaaaaa	36840
ttaattgcaa	ggacaacttt	tgaatctat	tattgaatac	ctaggacact	catgatgtct	36900
ctccacattc	ttgccaagga	attgctttat	gaattaaacc	acttggggaa	ttaaaccact	36960
atcaattaaa	catcacagttg	tataaggata	gacatatact	actttacatg	caaaaattta	37020
ggccatagca	atttagtaac	taccacgtta	cagggactgg	aacttttcac	tcctccactt	37080
cttctatctc	accactctgt	tatatgaagc	tatgattctc	cattcagaat	ataactacat	37140
tagtttttga	aataggaaaa	tatgaaggac	agactctaga	aatatcagaa	tcacttggac	37200
aacaaagtcc	aaaatatgta	ttatgaaagt	gaattatcat	gtttgactca	atatgttgct	37260
gcaaataaat	aaaaattatt	tgctcaatat	agatactcct	tatatgcctc	aatattttat	37320
atacactgac	tgaatattta	tgactatgta	tgagcagaat	aatgagaaaa	aaacttacta	37380
aaaaatctaa	ttatttttgt	cctttttctta	ataaaagtcc	cgtttcacag	ttataacact	37440
ttctggattt	ttattttttat	tctgacttta	taaaataaac	attattttta	gaatagtttt	37500
aaattaacaa	aaaaacacac	agagattaca	gaatgctcca	tatatgttat	acatatatat	37560
atatattata	tatacacaca	cacacatata	tttaaatacg	tatatatat	atatatttta	37620
tatatcccg	ccctccctcc	agaaatcact	gaattgttta	ctggctctaa	agttttgctc	37680
tcattattaa	atatgtaaca	aatatatatt	atatattttat	atacaatgta	tataaaataa	37740
atatatatatac	acatgtaaaa	aaatatatat	atattttgct	tgctccagaa	ttccattcca	37800
tccaggatac	cacagaactt	tcactaatct	tgtctctttg	ggtgcttctt	ggctataaca	37860
gtttcttagg	attttcttat	ttttgatgac	cttcacagtt	ttgaggagt	ctggtcagat	37920
attttgtaaa	atgttctctt	cttgtaattt	ttctgatttt	ttacctgat	tagactggac	37980
ttagggcttg	tggggatgga	ctgtcacaga	ggtaaaaatac	attctcatca	cagcatatca	38040
agggtttata	ctattgacat	gaacttaaac	tattgatgtt	aacactgatc	acctggttga	38100
ggtagtattt	gtcaggattc	tccactgtaa	agtttaatttt	ttccctgttt	ctatatgtga	38160
tttttagctat	atcaataaga	actcatggac	atttctcttg	tactttgggt	tataatccaa	38220
tcttttttat	ttttgctcaa	attgttccag	ctttggctcat	cggacacttt	aagtagctgc	38280
tgtgtgcttt	tgtatttatt	gcagcttgg	ctgttttagca	ctttcttact	ttctaaaatg	38340
tcaagatgct	ccaggcttat	cttgatatatt	accttatgtc	gctccaggat	taaccatttc	38400
tccaagggtt	tgacaaactc	atagtgtcat	gtattcacca	ttacagtatc	atacagaata	38460

atTTTtacagc	ccgagaacag	ccctgtgatt	caccagtgca	acctttccct	cccttagaaa	38520
ttactgattg	cttactggct	ctaaagtttt	gctctttaca	gaatatcaca	taaatgaact	38580
catgggtatg	tagcttttac	agactgaact	ccttcattta	gtaatatgga	tttaagattc	38640
atccatgtct	ttgtgtatgt	tgatagccca	tttttaaata	actgaataat	atttcacac	38700
atggatatac	catgggttat	ttgttcattc	acctattgaa	aggcatcttg	attgtttcca	38760
atTTgggaca	attataaata	aagttgccac	atgtataaat	atgcatgctg	ctttttgtgt	38820
ggatataaat	gttcaaataa	gttagttaaa	cacctggtag	tctgattgct	ggattatatg	38880
ataatgctac	atTTTtccct	gttaagaaac	tgacaaattg	tcttctaag	cagttgtgcc	38940
atTTtgagtt	ctgatcagca	gtggatgaaa	gttcccatta	ctccatttca	ttgtcaacaa	39000
ttatTTtcat	tatgtttgtt	tcttttttgt	gagttttaaa	atggctttgt	atattttata	39060
tataagtaat	ttattagata	tgtgttttgc	aaatatctcc	ttttagtctg	tggattgtct	39120
TTTTattcct	tgaacaattt	atTTTataaa	acataagttt	ttatTTtga	taaagtccaa	39180
cacatctatt	tgttttaattg	attgtgctat	ttgtgttatg	tttaaaaaact	catcaccaaa	39240
tccTgggtca	gacagatttt	ctattatgtt	ttgtctaga	aatttcacat	tttgtatttt	39300
acatacattt	gtgatccatt	ttgatttaat	tttgtgtaag	gtataagggtc	tttttctaag	39360
tttattttatt	tattttatttt	tgtatgtgga	tgttcaatta	atccagaatc	atttgttgaa	39420
aaaactattc	tctttttcct	tcaacctttt	tgacctttt	gtcaaagatt	aaatgatggt	39480
atTTctgtac	aataatcata	tgatgtggct	gtgtcccaac	ccaaatctca	tcttgaattg	39540
tagttcccat	gattcccacg	tgttgtggga	gggacctgt	gaaagataat	tgaatcatgg	39600
gtgaggTTTT	ccctatactg	tccTgtggta	gtgaataagt	ctcatgagat	ctgatggttt	39660
tataaggggc	acctctttca	ctcggttccc	atTTTctcct	ttgctgccgc	catgtgggac	39720
atgcctTTTT	ccttccacca	ggattgtggg	cctccccagc	cacgtggaac	tgtgagttcca	39780
ttgaacatct	TTTTctttgt	agattgcccc	gtttctatca	ggagcacaaa	ataaaactaat	39840
acogatgata	tctgtcTTTT	ctgttctggt	ccattgatct	aagtgtttat	tgattggcta	39900
atatcatgct	atcttgatta	ctgtagcttt	gcagtaaatc	atgaaggcta	ataggacaag	39960
tcttttgact	ttcttcttat	tcttttttac	tactgtattt	gtttttccag	gtcttgtgcc	40020
atTTgtatga	aatttagaat	cagtttggtt	ccatctacaa	aatagcttgc	cacgattttc	40080
TTTTtgTggt	tacattgaat	ctacagatta	ggttgaagga	aattgacatc	tcaacaatgt	40140
tatgtcttct	aatgcataaa	catgaaatat	tgtccatttt	attgagatct	ttgatttttt	40200
catcaatggt	ttatagtTTT	caacatacaa	atctgtgca	tattttgtta	gatttatatt	40260
tgagtatatc	atTTattgga	tgctattgta	aattataagg	TTTTttttac	atTTtaattc	40320
caaactcttg	tctttcatgt	atgagaaagc	catggacttt	ttatattgcc	ctgatatcct	40380
gtaaccttgc	tgtattcact	tattattcca	cctaagcgga	tctcagctgc	atctgctgga	40440
tgaggTggtg	ctttatccta	taacttctat	tctctgatgc	atccaataaa	attcattgat	40500
ttcacttttg	tctgtcTTTT	ccttgtaaga	acaggaaata	tgaagtccaa	gctttcacac	40560
ataggagata	aaactgaaag	tccacacatt	ctgattttta	aaaatcttct	attcttttgt	40620
tttgctattc	atatgtgttt	tttaaagtgt	tatgcttgag	gtttgaattg	tttatgtaat	40680
gtagaatcca	cagataaaca	caattagcaa	atgtaaagcta	taattcttac	agttttaaag	40740
gtacctttgt	atgtacccca	caaaaaatag	atgtcaacaa	tttgagggt	actttttatt	40800
taagtaaat	tacacctgct	ttctcttttt	tttaattgaa	cccagttgca	gttctttgaa	40860
caccagggaaa	aacttctctag	caacctctga	tgtattcctg	catctagagt	agaaacacag	40920
cacacagcac	agtagagaat	tgaaaaattt	gagaaaaaga	gtcagttgtt	ggccacattg	40980
actaagactc	aggcacgcac	atacacatac	caaagtgtt	agcaccttag	ttgagctcaa	41040
ataaggaggga	aaaaaaccag	cattacaaga	gtgagacagt	ggcaggttac	tactagccct	41100
gcctgtggta	ttctcttccc	atTTgaccag	aaatggcaaa	ttctgttgaa	aagaatggcc	41160
catTTtccagg	caaagctTTa	ggaagcacat	gatttccaca	tagaatctgg	cctttgtcag	41220
acatgggtatg	ggactgtaaa	aatgacttat	tccaagatga	gaacacatgg	ggtagagctt	41280
taaaaactgc	aagccatgcg	ttgttctctt	attgaagttt	tgttaaagct	gttccagaca	41340
tttgctgcct	ttgcttccat	ttttggactt	gagagagaag	aggcatcaag	acgtctgcag	41400
gggggacatg	aataagtgtg	tgtgcacatg	cgccctgtg	tggatatgtc	aaaataataa	41460
agTTtaataa	ttgatttcta	taattattaa	atatagcttt	actaattgat	agctataatt	41520
attaaacata	gctataatta	ttaaatatta	attattttat	atgtttagaa	aaaatcctat	41580
acctttgata	atattcatgt	gtgcaaaata	tattcttctg	tcttggtaac	tgctctgttc	41640
ataatcatct	TTTTttggac	aaacaaagaa	aacatataga	caaacaggat	taatataaaa	41700
tttaacaatt	atatagcaat	acagtggcta	actatagcta	ctctaattaa	ttacctgttc	41760
tagagtgatt	ggactatcgc	aggtataaac	octagcaggg	ataaaatctt	tttatttata	41820
tgctatataa	tttaaagtatg	ttctaaaaaca	cttgcatgat	ttattccaag	aatactgcat	41880
ggatttctgt	gcacgtgaga	atTTctctgc	ctgttcttcc	tattaaccac	taaatgttat	41940
gtcatgctgg	ggatgtgagc	cctgaagcag	tgacaggaaa	tatctgcaaa	atggcaaatg	42000
cctcccatta	ctgagagtgc	tgattacaaa	gctacaatta	catgtggatt	attcatgttt	42060
gcagcaaat	aatgaactgg	tttatagttt	ttcaagaaga	aagcatctgt	agtatagtaa	42120

aaattactct	gaatcttggg	tcaaagaatc	cggtttccaa	tccagccttt	gttaccagta	42180
gctgtgttag	ttaactattg	ctgtgtaaca	aattatccca	agacttagtg	gcttaaaaca	42240
gtaagtattt	gttatctctt	acagatttga	tgggtcaagc	atttgggaat	ggctaaactg	42300
gacggttctg	gcttgggagc	tttcatgagg	ttacagtgat	gatgccagtc	agggcagcag	42360
ccacctgtag	gtatgacagt	ggctgcagca	tcggcatccg	aggtatctta	tttatgtggt	42420
attggcagaa	gatatccact	cctcactggc	agcttgactg	gcctcacttt	ttagcagctg	42480
actttcccca	gaatgggaga	tacaagaaag	agagggtgag	aaggaaacca	taatgccttt	42540
ttgacctggg	cttagacgtc	acaagtcacc	atttatacac	attcagttta	ttggaagtga	42600
gtcactaagg	gtggtcatat	ttcagagcaa	gaaaactagc	ctcccttgta	tgaagtggag	42660
aatgatataa	aaaaaagggt	gtattttcat	accaccatgg	tggccatatg	agcttcagta	42720
gttaaacctc	gctgaacatc	agtcttctcc	acaaatgaaa	ataacaatat	ctacctcaaa	42780
gaattattgg	ggctgttcag	taagatgatg	catctgacat	attttgtaaa	atgtaagcta	42840
ttgcatcagt	gtaaagtgtg	ttgaatgtaa	tgctcagtgt	gttttgtcat	agacatcacc	42900
aatgcatggt	gaggaaatct	tgtttagtga	cttccaagtt	aggcagacac	tggtactgtg	42960
cacctctctc	tcaaaaaaga	agcatgcaaa	tctttgactg	ttaaagtata	tagcattcaa	43020
agattttgtt	gtaatcattg	agaataataa	tttaattatt	tatatataat	taattattgt	43080
ttaattatta	attaaattat	tatttaatta	tataataata	tataattatt	taattattat	43140
ataatgaatc	aggtgcagtg	gttcacgcct	gtaatcctaa	tcccagcact	ttgggaggcc	43200
gaggtgggtg	gatctcctga	cgtcaggagt	tcgagaccag	cctggccaac	atggtaaaaa	43260
ccggtttcta	ctaaaaatat	aaaaattagc	caagcatggg	gggtgcagtc	tgtaatccca	43320
gcaactcagg	aggctgagac	aggagaatog	cttgaacccc	aggaagcgga	ggttgcagtg	43380
agctgacatt	gcaccactac	actccaagtct	gggcaacaaa	gcaagactcc	atctcataaa	43440
taataataat	aatttgttaa	tgaaaaatgt	taatttccaa	agaatgatag	agttttaaata	43500
tcagaattat	ttatcttatt	atttttgggt	gattgcttaa	atatacagac	tttgtttttc	43560
gtgtttttgt	ttgttttccc	atatgacaag	aggtcggaga	ctggcaggca	ggcagcatga	43620
gcttctgcag	cagcgcagtt	ctgtgagcca	tggccctctt	gtttctgggt	aatcaccttc	43680
gtccttggtg	ccttctgcct	gctccccttg	cagcactgat	ccttcattcc	aggcagatgg	43740
aataagatca	aagcacaaaa	aaaaaagttt	gctgaatatg	ccctgtttta	ttttaatttc	43800
ttattttcaa	ataaatttag	acttttggaa	aaactgtaaa	gatagtagaa	agtttttcta	43860
tacttttcat	ccagattcac	ttaaggataa	cgtcttttat	aatcatggta	tatttattga	43920
aatcagaaaa	ccaacattca	ctatgagcca	attgggtatta	gttacgctat	aactaattat	43980
tattaactaa	tagtatcatt	atttacttat	ccacaataca	atgatgtaat	aacaaaatgt	44040
agtatataat	attaaataca	gcaatgtagg	atagtaatga	tagtactact	gctagtttta	44100
gttaataata	ttaactaatt	tacaggcctt	attaaaaat	tttcagtttt	accactaatg	44160
tctttcttct	ggttccaaga	ttcacaaatg	atttatgtgt	catgtcacca	ctgttctctg	44220
caatctgtgc	cagtcctcca	ctcttttctt	gtcttttgcg	aacttgaaac	tcattggagaa	44280
cgcttgggtc	gttattgtgt	atagaatgcc	cctcagtttg	ggttagtctg	atgttttcac	44340
atgactaaat	tgagggttatg	cagtttttca	aaactgcggc	agatgtgatg	atactgtgct	44400
tttccactat	atgatagcat	gtgatgtatg	agatgtaaac	acggatcatt	gttttctttt	44460
cctttttttt	tttaagacgg	agtctcgctc	tgtcccagg	ccagagttca	gtggcatgat	44520
ctcggtcac	tttaacctct	acctcccagg	ttcaagcgat	tctcctgcct	cagccttcca	44580
agtagctggg	attacaagca	tgcgccacca	ccccagcta	atttttgcac	tttttagtagg	44640
acgagatttc	accacgttgg	ccagaatggg	ctggatctct	caacctcgtg	atccaccac	44700
ctcggccttc	caaagtgcgt	ggattatagg	catgagccac	catgctcagt	ggatcatttt	44760
ttttttatgt	tttgtttgtt	tgtttgtttt	gattttctatg	ttgtctgtca	gatgtctttt	44820
aagtgatttt	tttttaaat	tataacttgt	aagtgtcttc	tgggaagata	tttttaggct	44880
atggaaatac	tctgtttttc	atcatacact	tgcccactaa	tttttagcatc	aataatggct	44940
tttgcttgca	gtgactactt	atgtagtgtt	taatgtgatg	atttctgtag	ggatatataa	45000
atgagccaat	gagaaaccac	agacatcaca	tttctagctt	aattctgaaa	ggtaaacgtc	45060
tcatttggcaa	atttttaaaa	gttaagacaa	tttggtttca	tcaggagacc	caaagaaaaa	45120
atgagtagtt	tgaatttctg	gtcataagga	tagtattctt	gatcttctta	atatggaagc	45180
aactaaatat	atttacaaca	ggaagaacta	gatatacttg	caaaattcaa	agagaaaacc	45240
aaagcaatac	agccaagtct	tatttgcagg	tcgagatctt	ccaaagaaat	ggcagtattg	45300
tgatagtatt	gtacttttaa	aactggaaag	agatgggaaa	ggtacaagtg	gatgttcatc	45360
atggcgtaga	catcaatata	tgatcagtc	cttgttttaa	tgagaagaaa	45420	
aagctgtgat	tctttattta	ttcaaaatca	ccatcacttt	tttcttttaa	atagtcagac	45480
tactagggcc	tctgatactt	attggaagat	aaagttaagg	agagaaaagg	ggtgtataaa	45540
gtaaaaatat	atggcagctg	tccaacaaag	tcaataatcc	acttttagga	ccccaaaaata	45600
gatttttctg	agtgtctgtc	tgtgcacaag	tttgtcagca	tttgagagaa	atgactgagc	45660
agccacaaca	aagagtgttt	aaacaaacca	atctggatgt	caaaagggtta	tgtgccgcag	45720
agggctgctg	gtgttaattc	tatatggatt	aacacacag	cogtttctta	aaatagaaaa	45780

acttgaaaga	agtgaagtga	ctcagagacc	tttttttttt	gtaggaagtc	tagatgagac	45840
atttattttt	ttctttgaac	tggtgaagtt	ctctctactc	agtgcgactg	gtgaggttgc	45900
atgaaaagtc	tgtgaatcaa	tcacaggacc	cagttctgcc	ctgagacctc	aaggagagag	45960
gaggctatgc	tggaccccg	actgtagagg	atccccactc	ctacctctt	gtggacaatg	46020
caccctcgg	catggggcaa	gaagtccaag	aacgccataa	gccaagggaa	cgtacacaca	46080
attagcctgt	gtctccttct	gcacaatgct	ttctgcggct	cttcttgc	gagtgtggag	46140
ctcatttgca	agagcagatc	tcttgcttgg	attcatgttc	ttgtggtcac	accacacccat	46200
ccccagtatg	ggcctctctc	cagcaagggg	atgaggatgt	tggttgatcg	tggggctaatt	46260
gtacatgacc	actgggaccc	tcgaaggtgt	ggcaggctag	gcaggaatag	cagagaattc	46320
actacagggg	tgtgtgtgtg	tggtgtgtgt	tggtgtgtgt	tggtgtgtag	taaggaaat	46380
agagcaggag	gctttcacaa	tattccagat	ggaagacagt	gaaatccgcc	aagcctaaat	46440
tcaaatcata	ccaaaaggt	taaggagtta	agagaagaga	ataggatata	ccttacttgt	46500
ttacaaatt	gtttcctttt	ttatataaag	taatttttag	tggtaaaaa	tacagaaaat	46560
ataaatatat	acaaaacct	cataaaattt	accattttca	tcatttttca	gtgtgaaatt	46620
cagtggcttt	aagttcattc	atgatgttgt	acaaccatcg	gtactatata	cagaacattt	46680
tcataatttt	aaacgaaaca	ttgttcccta	ttaaacgcta	acttcccaat	cccctgtacc	46740
ctcaggcctt	ggtaaccatt	attcttccct	ctgctttcct	gcatttgcct	attctagatc	46800
ctgaatatag	tggacacatg	cagtgttttc	cctttatgtc	tagcttcttt	cctgtagcac	46860
agtgttatca	agtttcatcc	atgtgttgtc	gtgtatcagt	acttcattca	cataatttta	46920
tagataaata	atatttcatt	gcagtgtgtg	accgcataat	gcttatccat	tcattagtgt	46980
ctgggcactt	gacttgcttc	catgttttaa	ctcagtgaa	tatgctgcta	tgaacattgt	47040
tgtacaagtg	tctgcttgag	tcctgtttt	caattttttg	gggtatataa	ttaagagtgg	47100
aattgctgga	tcagacagaa	actgtatgtt	taacattttg	aagaactgac	aagttatcgt	47160
ttccttggtc	ttcccttata	ctatttgca	atgtagcatg	cagatctcca	ctgtagccct	47220
gggtcctgca	aataattatta	gggacaggac	gtgacagcaa	ccttatggat	ctatacatgt	47280
agacagttaa	tcaatccttg	taaatgttag	tctttcactt	ttttcactt	ttgctttcca	47340
tattatataa	atgagaaatc	tgggtttcca	tcagtgttca	aacttgaaa	cttttgtgtc	47400
atattaaaaa	aaaaatccaa	atcaagctgt	tttgtggtat	ttcagatctg	actgaaacca	47460
ggaagtaata	gaaatataaa	aagaaaactt	gatcttctga	ggttttcttt	accaaagtta	47520
ctgggataaa	aattaagcct	ttaaaatttt	tgaggttcat	atatcaatgc	cactgatttc	47580
taaaaaatat	tgtattgact	gcttcttttg	gaaggtaaaa	aaactcagtt	ctcctacaag	47640
actttgtaga	aattcctttg	cttaaatcat	atagcttaaa	aaatgagcct	tttaaaactg	47700
tgaaggatat	cagaaagaaa	tgtgaagaaa	gaagaaaagc	acaaattaaa	agaacgaaca	47760
agagctacgt	tagcagtgag	acccaattta	aagagtttgg	gtccagccag	agcttttcaa	47820
atcttttgaa	gottatacag	tgttctgata	cctaacaacc	catattcttg	aaaaattaac	47880
cagaaaaact	gcatgagcta	aatgtcttta	tgtttgaatt	agagaacagg	gaaaggtcct	47940
gctgccttta	ggaaaagcct	ttaatcagaa	tgtctcttca	tcctggcttt	tgacctttcc	48000
taacattcaa	ttgttttttc	acactgctct	tgaagagacc	tgccgctcta	agaaagaagg	48060
atggatatct	tccccttgct	tagcacggtt	tatgtcaaga	ttttgccctt	gtaaaagcac	48120
ctaaatgtgc	tggtgacaga	cagagaggga	tgatcattca	aaagtagtct	ctgcttttgt	48180
taaaccttta	tttcaaaaaca	atggagagca	catcccatac	ctgctacgaa	aagttctaac	48240
ttccaattca	taggtcattt	ttttgaaatt	gtattttctg	attaaaaaca	catattttct	48300
caactagtat	ccatccttat	ttctagaggt	atcacctgtc	atttcgcacc	ccctaaataa	48360
ccttcttgcc	cogtaaaaaa	tattttgtaa	gtattcgtat	gtaccacag	tcataatta	48420
ggacatctca	cttctctttg	ttaaaaattc	cttatgttta	cagcttttca	aagggtcttg	48480
cattcactaa	ctaagtagtc	ggaaactttt	gagacagctt	tgcagaacaa	gacacaacaa	48540
ctacatgttg	agtaacttgc	tcacggttac	atgatagtta	tgtaagatgc	ttggtgatag	48600
gactgggcgc	taacatgttt	ttccaagaa	aaataaacct	ccattgcac	tagtccacac	48660
taggacgatg	ttttgggcat	atgtttttcc	aagcatctgg	tagggacaga	tcaggtctaa	48720
gggatttctg	aaagggttaa	attaaatagg	atttcccttt	aaaattttga	tagagaaatt	48780
tagaaaaaaa	aggaatcaaa	tgotaattca	gaaggatttg	agaacaggaa	aaatattagc	48840
ttatcaggga	acttcttaag	gtcagagttt	cactatcaca	ttctagcttc	ttcagtgctt	48900
cctcattttg	gaaagtgttc	aggactttta	ggtgcagttc	tggacacctc	acatgtctga	48960
acatatccag	tggtctttct	taaagattat	aaggacttag	cagcaaatta	gtctgttttt	49020
cacatggctg	gagatggctg	aggcttaaga	atgctgcaca	gcacctctg	gtgattatta	49080
gtactcagaa	ggcctcccca	cctctcttgc	acctgtgatc	ctcatcttgc	ccttttccaa	49140
atatcgagca	gccaccataa	tatgtgtgtg	tgtgattaat	caattattac	taaaggatca	49200
gtaatcctta	acctctgaga	ttctcactgg	tatttaagga	ttttaaaaat	aggagaataa	49260
tcaatttaaa	cctttaagct	gttttgcttc	tgatactatt	ccatgtctta	gcattcaatg	49320
tatttttatt	ggtggatcca	tatgacaaag	tgaaaaggag	gagggttttg	gaaaaagaga	49380
tacagaagtt	cacattttag	ctaccgcata	actgtaaaaa	caagggcaat	tctgtattgg	49440



gcttacattt	atcttttcag	tgatctatt	taatagaaat	agaaataaaa	cacatatcaa	49500
taattgcacc	aagatactga	aggtagttt	tattgaaaa	ttgatcacta	aggaattgac	49560
agaattgaga	agtgaagaaa	tttcaaagat	acacatgcct	gaaagcagaa	atcatatcct	49620
atatgtcaaa	catgaaaatc	aaagtcagct	gcttgaatg	tgggccta	aaatcttgaa	49680
tctctgcaac	tgtagggaat	agttattgtt	ttctaatact	gcacacattt	tagtaaactt	49740
ttctgtcctt	aatgttaaac	ttcatatttc	ctacacatct	tttttaagt	taattatttt	49800
aatatgcatt	catatattta	ccacttatta	catgggaaga	aacaattctt	aagtctgaga	49860
aggttggtaa	tcattttacat	ttctgacaca	aaactgaagg	catggttgag	aattagaata	49920
attaaaaatag	cactggggga	caacctgaaa	actgtgactc	cattctgata	gaaggagtca	49980
ttgtccttga	acaatttgca	aatatgttcc	cttcttttcc	caaataatgt	cccttctttt	50040
cccataaag	gcttccatgc	tttatgcaac	taatgcgtta	tgaattgcag	ctctgtttct	50100
tacagagaag	tgaatttaca	aaagacatac	ctttcacatc	atgttcacct	ttttcacctc	50160
aatgacaccc	tgctcccttc	tcttaccttc	cctgacacca	cacagttcac	catcagctct	50220
tcctccttct	caacacccta	acttctaacc	tgcaactgaa	gctcacagtt	gtgcttagta	50280
tctgctgac	agttcccttc	acttgacaag	ttttaaaagt	caactttctt	tgtcttacag	50340
tttattgtct	tcagtctttt	atttccctta	ccttttaata	aaaggattca	gtattccatt	50400
tctattaatc	ttttattaaa	gtttattaga	aatagaatgc	tgaatagaaa	atagggaagt	50460
tagctgtagt	taagattgtg	attattcatt	atactgatat	ttttcttata	ttgtttttga	50520
actgatttat	gggtagttag	ctccaaaact	gaaggaaatac	tatgatattt	tcaacactat	50580
taccagcaact	atgtcaggta	tttaagttta	tccatgagca	gagtgcatt	tcaagatcat	50640
gaattttcac	aaacttcaat	gttactactg	tgtttatgct	gaataaatgg	tttctttggc	50700
acctctttca	cacacacaca	cgcacacatc	tatacatggt	aacataggct	aaagcaatta	50760
gagaagttca	ggtatttcaa	tgactgaatc	acaaaaatat	gttattgatc	tcatttgtaa	50820
atcaatgaat	ggttaataag	aagattaaga	agacaaatta	caaggtcaca	gccattcag	50880
ctgaagcatg	ttgttatggc	cagaggctca	tgagaggaaa	aggtgtggct	aagaacagaa	50940
gatccgttaa	gctccaagcc	tgttcaatta	tttaagtcac	tgagacaaag	tagatgtgga	51000
gtgtgagacg	ccagtgtaga	gccttcaact	gattttcctt	ggggaggttg	acctgactta	51060
atagtgtggg	aaatcagggtg	aatttttggt	gtgactagat	ttgaccaaga	tggaagtttc	51120
catgaaaatg	caaaccaatc	tttgaattt	gaattccata	aagcaaaaaa	ttctttacata	51180
aatgcttata	attttaccac	tctaccactg	tggtatgcac	gccacgagat	ttgtgtagta	51240
gttctaaaaa	gtatgtcttc	ttggacaaca	tccaaagtag	accacatgca	cactgaaaca	51300
agccttacac	aatagctttt	ctatatatgt	atctatatct	acatctatat	ttatatctgt	51360
atctagctaa	aagaaaagct	ccacatactg	atagaattat	ataaaagcaa	atactaactg	51420
ttgggaattt	cgattataat	acacttaatt	gaaaacaaaa	caaaaatatt	ttttaaaagta	51480
ttaagcaagc	atagagagac	attttggtgc	ttacaaggat	ttctgggaat	acaagcaatt	51540
attttttgtc	agggaccata	tatttaggta	cagatcatta	acaatgcttt	ctttaatatt	51600
tatttagtga	attgtacatg	tttaactctaa	gggcaattaa	aaccataat	tataatatga	51660
gaaacttttt	ctccacttag	tattttttat	tataaatatg	gctagcattg	aatcattggc	51720
tataaccatg	taattttaat	gcttgggaaa	catctgttag	aaaaactttc	cctatagaag	51780
acatttatca	atgtcctgcc	ttatctctgg	ttatctttgg	tttattacat	tttatgataa	51840
ccataaaaact	gaaaccatcc	tcattttacat	ttttcctctt	ctaaatttgg	ggaaattata	51900
ttttaagtgt	ttatcattat	ctcattatat	aatctacaga	aataataaac	agcagcatac	51960
ctgattttta	aatgagactt	ttgaatttct	acaataaata	gaatataacg	aactctgctt	52020
ctttgggtat	atcacgtcaa	atatacatta	atgagcagat	agctttaaca	tgtttatcat	52080
gttataatct	gattgttgtc	tttaactaat	aagataccta	atttaactga	agtatcaagt	52140
gatctcaaaa	cctcagtagc	atctgataaa	ggttgctgtc	tcacctgtgc	cacatggcca	52200
ttgocatttg	actgctgctt	caccacatc	acctgcactc	ccaaactcag	cctggctgga	52260
tatcttttta	caagatacct	gataatgaag	gagagaagga	acatggtgaa	cactatgcag	52320
tctcagaaga	tctgcttgga	agagacatgt	tacttctgct	gtcatttcct	tgccagagc	52380
aagtgaatg	ctgaacctga	tcagtggctt	attgtttgaa	cagtaataca	acctactgtt	52440
ctatcataga	atgatattgg	agtcataat	ttcaattagc	cattatcaat	gtctctcata	52500
attttcttta	cacattttct	cagaatctag	agtaaaattt	aaaaaaggaa	agacagataa	52560
atgctctcat	ctccaaattc	agtgtcctaa	attccactta	ctttaatcag	tataaaaaatc	52620
aggtagctca	gatttttaag	gactagtcac	ttttgtctct	ttttttttgc	tatcagcgtt	52680
tatatgtatt	atttgggaaga	aaaaagagcc	ccctcaacct	tccagagogg	acacacacac	52740
gtacacacac	acatcggaac	cattagagag	gctgacctga	aagtgtttta	aaaactatga	52800
cacagttaaa	ctatgatttc	aaagttttcc	caagcaagtt	atatgtcaga	atatgagtag	52860
gggaaaacta	ttgaactgcc	acaattctgc	tggtgggttc	attggcaatt	ccaattatgt	52920
tctgatcata	gtgtcatcag	agcaattagt	tcatcccaaa	cttgttttgt	attgccactc	52980
cacagggaaa	atagtaaatc	tgctcactg	actgcagttc	tctaaaatgt	ccttaaagtt	53040
tagttatttc	aaacttagtg	gaagaaatat	gttggttgta	gtggtcatcc	ttgctagtag	53100



tctttaatca	ttacctgcca	ggaaacactt	tttattttacc	agtgcacatgag	agtttaggtg	53160
agaattattc	tactaccgaa	agctcaagat	gtattttgtgc	tatagttata	aatatagatt	53220
ttttttttgt	aattctaact	tttcaggtag	ttttctcttt	agctgtttta	tttttagtata	53280
tgtatttttt	ccatttttta	tggaattgta	aggcatattt	ggtggttcta	gtttacaact	53340
ataacatcat	ttccagagtc	atcaattcct	ttgtcaggac	atgacattat	gtttttccat	53400
tttttagact	tcctaaaatg	tcttgataac	caaaacacag	aggggttatca	ttatacagtt	53460
atgtatgata	cggtttcctc	tatataatgt	gtcacataga	gtattttatt	tocagccaaa	53520
gatcattatt	ttttaaagaa	actcaggaac	aagaaataag	gatgcactca	cacaaataga	53580
cagtagtgaa	ttcagaatgc	atagcagagt	gtgcaaagta	aagcaaagtc	atccatgccc	53640
aattaccta	tttgtaattg	cagctattca	ttgcacttga	ttctttcctt	caactccttc	53700
ctctccttcc	ctccccggag	ctctccttcc	ccagcggttg	tatggcagct	tcgtaataaa	53760
cccagaacac	ttagcaattt	ctgctccgat	ttgctctagc	ttcaaaggtc	tatggctcatg	53820
attacatgca	ataacgatac	attcaatgtg	aaactacaat	ctgtgtaact	ttgtatttct	53880
ctttgaaggt	ttagatacga	ttgactgtgt	caaccataag	cagaataata	attcctgtgc	53940
tcagagacaa	gtcagagaga	gagagagaga	gattgacaaa	caaacaaaqa	aacaaacaaa	54000
caaacaaaaa	acagatttta	agaactgagc	cctaattcat	ttttttttct	tttttattgt	54060
ggcaaaaatac	acatgacatt	tactatctta	accattttta	aatgaacagt	ttagtatat	54120
taaatacata	tattacattg	tgtagacatc	accatcctcc	atcttcaggg	ttctttccac	54180
ctttcaaaac	tgaaattcta	cccattaaac	aattcctcat	tccccccctt	ccccagctc	54240
ggctgctggc	aaacagttct	acttattgtc	tgtatgattt	tgtacactgt	aagtacctca	54300
tagaagtgtg	atcatcaga	atttaccttt	ttgtgatggg	cttatttcac	ttagcatcct	54360
aaaggtttgt	ccatgttgta	gtctatgtca	gaactggctt	cccttttaag	gctgaatgat	54420
attccattgt	atatataaat	cacattatat	atgttatatt	ctaagcattt	tatatattta	54480
ggccttacat	ttgggtcttt	gattcatatt	cagttaattt	ttatatggag	gttcatataa	54540
ggttcaactt	cattccgttg	catgtggaga	ttcagttttc	tcagcatctg	ttgttaaaaa	54600
tagtgccttt	ttcccattaa	atgatcttgc	cacccttgct	aaaaataatt	tggccatgta	54660
agtattttat	tctattggcc	aacatgtttg	ttttacgcca	gctgatataa	ccatatgtaa	54720
ttattcactt	attagttctg	aattcattta	ttagtttcat	atgggatttt	ttgggtggaat	54780
cataagagtt	gtctacatat	atgattatat	catctatgga	cagagataag	tttacttctt	54840
cctttattcc	aacatatggt	ctctcttttt	cttgccataat	tgtcttagct	aggatttcca	54900
gtactacggt	aaatagaaat	gaggaaagca	gacacccttg	ctttgttccct	gatctgacag	54960
gaaaaacttt	ttaaaaaatt	gtgatgaggt	agtttccttc	tatcctagta	tactgagtat	55020
tgttttaatt	atgaaaggtt	attgaatttt	ctcaagtgtc	tttactgcat	caataaaaaga	55080
tgattacaatg	tcttttttct	tcatttagtgt	ggtgtattcc	attgttaatg	tgggtatattc	55140
cactgtttta	tttttgtgtg	ttgaaccatc	tttgatttcc	aggcataaat	cctactttga	55200
ttttcgtagt	ttgaaccatc	tttgatttcc	aggcataaat	cctacttaga	aatggtgtat	55260
aattccttta	atatgttgca	aaattcagtt	tgtctagtag	tttttgaaga	ttgttcccat	55320
taatgttcat	acggaatatt	gatcttttag	tttctatttt	ttgtgatgtc	ttgtctgac	55380
ttgggtatca	aggtagttct	ggcctcatag	aatgtgttag	gaagtgttcc	ctcttctttg	55440
atattttgga	aattattgaa	aattggggtg	agcccttctt	taaatgtttg	gtagaattca	55500
gcaaaaagc	catcaggcca	aggacttttg	tttgtcagga	tatttttgat	tagtaattca	55560
tctcctgact	cttaaaaatg	ctattacaat	tttctatttg	tttgtggttt	agattttagta	55620
ggctttgcgt	ttctagaaat	ttgtccactt	catccagggt	attcaatttg	ttggcatatc	55680
attactttcg	taataactctc	tgtctgatctt	tgtatttatg	tagacttgga	agtaatgttc	55740
ctatgttccct	tttttttttt	tttatttttg	taacttgagt	ctttcttctt	tctttctagc	55800
taaaggtttt	ctaaatttgt	tactattttg	aagaaccaac	ttttggtttc	atttattttt	55860
cctattattt	tacattctct	attttatctc	agctgtaata	tttattattt	catttttttt	55920
gctagtttctg	ggttcagtat	attcttattt	ttttagttcc	ttaagttgta	aagtttaggtt	55980
gatgatttaa	gatctttctt	gttttttgta	atgtatgcat	ctaaattttc	taacagtgtc	56040
ttcattatgt	cccatatggt	ttgacatatt	gtttgttcat	tttoggttat	ctccaagtat	56100
tttctaattt	ctcatgtaat	ttattatttg	atccattggg	aagagtatgt	tgtttaattt	56160
ccacaaattt	gtaaaagttc	tcgtttttct	tctgttactg	atttcttact	ttatoccat	56220
ttgatcagaa	aaatgttttg	tatgatattt	tttaccgtaa	gttttttgag	acttaatttg	56280
tgacctata	tatgttctat	ccaggaaaat	gttccatgtg	tacttaaaat	aatgtgtatg	56340
ccatagaaac	tgggtagagt	gttccgtatc	tgtccattag	acctagtgg	tgtattgtgc	56400
cattcaagtt	ctctgttgcc	ttacttatat	tctgtctcat	tgttctattc	attattgaga	56460
gtggggtatg	ataactacaa	caattattat	agaactgtat	gtttctgttt	tcaattctgt	56520
cactttttgc	ttcatatatt	ttatggcttg	tttattagatg	cataaatatt	tataattttt	56580
atatcttctt	gtacatttaa	accttttaat	aatatataat	gtccttcttg	tatgttttaa	56640
aatctacttt	gtctgatatt	agtatagcca	ctcttgcctt	tcttcactta	ctatttggac	56700
aaaaatatatt	ttttcatcct	ttaactttca	gtgtatctgt	gtctttgggt	ctaaagttag	56760

tctattgtag	actgggccta	catcattttt	taaaatccat	tttactaacc	tctgtcttct	56820
gattggagag	cttaattccat	ttacacttaa	aatcattgct	gataagcaag	caattttgct	56880
atttggtttt	tatatgcctt	atagctttct	catacctcat	tttctacatt	tctgtgtttt	56940
gtgtttagtt	catctagttg	atttttcagt	taaatattaa	aagtcttttt	ttctgtctatt	57000
ttctttcttc	tatattcttt	gttggtacca	tggtgattat	attcaacatg	ctaaagttat	57060
aacattctaa	tttgaatttg	tgccaactta	attttaataa	tacactcaaa	tattttttac	57120
agctctgtcc	ccacctcttt	cagctgttca	tgctataaaa	ttatatctta	tacatttcat	57180
gttcgaaaac	ataaaactaat	aattctttga	tgcatcctac	tctcaaatta	tgtagaaaac	57240
aaaatgtgga	attgcaaacc	aaaattacaa	taccaacttt	tagcataata	attatcttaa	57300
tggttttagtg	tcttaaatga	agtagaaaac	aaaatgaagt	tatacagcgt	gttacaataa	57360
tactagcttt	taaaattgcc	tatttttaat	tcagatttaa	tttttgata	gcttttagtt	57420
attgtctagg	gtcctttaat	ttcatcccaa	ataactcccc	tgagcatttc	ttgcagggaa	57480
ggcttagagg	taataaaactc	ccccagcttt	tgtttaacctg	gaatgcctta	atttctccct	57540
catttttagat	aacagttttg	ctgaatctag	tattgttggt	tgacagtatt	ttacacttag	57600
tattttgtat	gccttctgat	tttcaactga	aaactgagca	tttgaatcta	ataatatgtt	57660
aactctgata	accggattct	ctttctttct	gaggggttgc	tggtttgttc	ttgacttttt	57720
tttattgttg	cagctgtctc	tgtgccaaagt	attagcatga	gggtgaaact	tattctctcc	57780
tcaggttttt	ttctgaactt	gtatcttccc	actgtgtcta	catggtgact	tcctaatttc	57840
acctatata	acagttgctt	gtgaatgtct	ttatcttcaa	tgtcagccgt	ccaaaagagt	57900
aaaaagagaa	aaacaaaggg	agaaaaaatg	aggcaccacc	ctctttacat	tccctgggaa	57960
ccacctcagc	caggggagag	gggctgacaa	taattgagga	ggagcataac	aatggccact	58020
gtttctttgt	ctgcacctct	gggatcagaa	gcagtgatca	gtgatcagag	cacgggtctc	58080
tggtatctgg	caggcagggg	tggttggtgc	ccaccctagt	ttctgcaagc	tggtgtgtag	58140
ctgctccagg	caccgctgcc	tgccacagag	tgagaggtg	gggaatgaat	agggtctatt	58200
gtgctaata	ctaaaattga	ccacaattaa	ccacaactga	cagtctaagc	catctcctgg	58260
gagtttgaag	tgcaagtctt	tccctggaac	ttaatttcaa	tagtttccag	aattccaaaa	58320
tagtaacatc	agacagattc	tgtcaggcca	ttattttttt	ctatgtaggg	atacatattt	58380
ctagtgtctac	ctactccacc	atctttctag	aatatttctc	tccccaataa	tttatatcat	58440
catttctggg	gtttgaagtc	aacaatctgt	attttaacaa	tttctctagg	tgtctctttg	58500
cctgtcttaa	gttttcaatc	cagttagccc	tctctatcct	caatataaaa	caaaagcacc	58560
atgaaaacat	ttggctctaa	gtgtgggtgc	catttaattt	atgttccaaa	ccagaaaaaa	58620
aaaaattgag	agtaaaaagaa	cgtgctatta	ataattcatc	tgtaataata	tatatgaacc	58680
tggactgtct	gaaacagttt	tttgagagac	acttagaaaa	aattatgctt	tggaaatttt	58740
ttatgggggtg	tcattaattt	gtttttgtcc	aatttttgca	acaaaatttt	gacttctata	58800
gtgaggtgac	attcataaat	ctgagtattc	atatctaaaa	acccaaaagc	atttttttct	58860
aaggatgtca	tgtagatcat	ataaaataac	tatagaaata	ttgaatccat	aatatgtcat	58920
atgagttgca	gccattttta	tattttgtcc	atttttatat	attctatgca	tataatatat	58980
ggtagttttc	tgatagtaat	ttccaattcc	tctcccaaaa	tggactggac	agatttctta	59040
gtgaggggaag	actgttttta	tatcattcaa	attttatcag	tccttttctg	tgggtccctt	59100
gctgttctta	ttccatagca	tcttatgtag	gattgatctc	ttccttggtg	tagagtccat	59160
attattattt	tcagaagtct	tggtttttcc	attggcaaat	gttcaaataa	aactttataa	59220
ccaaatcttt	ataatgctgt	tgccaacttt	gtgtctttta	ataaagtcca	attgccatag	59280
ggcaattgtg	cagtgggatg	gatatgtgaa	tctgtgtgaa	aaatgatatt	ttaaatgctt	59340
tgtgggggaag	agtaaccagt	ctggattgaa	ttgaatgcta	aattggcttt	acatttgga	59400
aaaattttaga	tggttatagta	gttcttata	aagaacatta	aacatctcac	tctaaatgaa	59460
gtacctcagg	tgatactgtg	atagctataa	ttgtttatga	cggaaatatt	cactcttaga	59520
agacttgggt	attgacaaga	acttggaaaa	ggatcaagca	gcaacatata	tgaaactgtg	59580
cttcagcagg	ggaaatacag	ttatgttcc	gggtcactgg	actcttggtta	tggactgaac	59640
aatccaagct	atataaagag	acataataac	tatgatcaga	actttgtgaa	atatgaccgt	59700
tctatttgaa	ttcaggttct	aagcagtatt	ccccaaagag	atctgatatt	gaatcaatgt	59760
tgtagctacc	agccttgaag	ttacctgaac	tttggtattc	tattattccc	ttagcttgta	59820
gaaactgact	atggtgaatt	cttaccocaa	atgtattttt	ccctagtaga	tgaactataa	59880
tctttcttat	tcatagtcta	tgattgggtgc	atcttctcta	gctccctttt	ttagatgtgc	59940
tgaggaagtc	cgctgtggtc	ctaaagctgc	acctgtgtgg	ctgttgtcta	atgtcccgtt	60000
ttggtgtctga	atgtcoatgag	tactaagatg	attctgagta	agagtccctaa	ttaatgccac	60060
caaaaatttgg	cgaaccccttg	tcttttgggg	aaagggggca	gaattcttca	cgatatgctt	60120
caaacctggc	ttcttcactt	ctgtgtacga	gataatttca	aactgctttt	ctcactttta	60180
aattttattc	atagatgtaa	aattttggca	ttttggctaa	tttaatttca	aatttgtatt	60240
tgcttgttcc	tgctttgatg	cgctacttat	gttgtgtttg	aaacatagta	gaaaagaggg	60300
tgctgcggac	aagtatgaga	agacatgctt	acagtcacat	aaaaggagaa	taaaaatatg	60360
tgtgtgactg	tagacogtac	ctccatgaac	attacaaatt	tctgactgcc	atctagaaca	60420

cgtgctttat	caacatgata	tcaatgatgt	ggatatttta	cttactttct	attttgttat	60480
acatgtttgg	tttagatgtg	acctttgacc	atttctgtat	aattttaata	aaatagattt	60540
aactacacaa	aaatcaccct	aaaagtaatt	ttccatgatg	catttttaaag	gtcatgtttt	60600
ccctgatgac	agcatgtttt	tgtagttaga	ccttgggaaa	gacaacaagg	agatctgaga	60660
gctaattgcaa	aattatcttt	actagctgga	tctccttgga	ctgggtcattt	atcctctgaa	60720
tcctaatttg	ccccttagca	ctcacaaatg	tcatctcccc	aaagccatag	ctgagcagaa	60780
agtggagctt	agcatttttt	attctgatag	aagagactgt	gtccatgcaa	ttatgacagt	60840
aaaaataaat	gagcatagct	tcaaacattt	ctttaataga	aacttagcct	ttgggtttaa	60900
ggttatataa	agtgaagga	atltgggtaa	ttaatatgaa	gtcaaacatt	aaaagattcc	60960
tcaagtaatc	caaaggtatc	aatgtttcaa	tgacataaaa	tacctctcta	gtccttatga	61020
tgagaagaga	aggtcctgat	gagtggctaa	gtagtgtatg	aagtotaagg	ctaaatttct	61080
atgaagaaat	attgtaataa	aatttttgaa	agatactcag	agatatttct	tgagatcatt	61140
atcatgtatt	ctgtacacaa	atgtattctt	tccattttat	tttcagtgtat	tttccagtga	61200
cagctataaac	tttagaaaagg	ctgattgtta	ttctttggct	tttctccctc	aattaaatca	61260
tgtttttttg	gtggaaagag	agatttaaaa	taaaatggaa	gcaagtggta	ttaatcaga	61320
gttactaaaa	ctgaaaatta	gggaaatgta	gaaaatataa	actaatatca	agctaattct	61380
tttctgtatc	agtgttagac	cttttatggg	tgaatgaggt	taaatgtgtg	ttagtataaa	61440
ggagttaaaag	gagtaagcaa	agacactaag	accagagcat	gactttatat	ctatatcagc	61500
ttgagtctat	ttatagttaga	atcactagaa	agtgttgacc	acacagagtt	gcattttttt	61560
cttttgatat	gcatacattt	tccctgaaag	aaaaaaggaa	ggaaggaagg	aaggaaggaa	61620
ggaaggaagg	aaggaaggaa	ggaaggaagg	aaagatggaa	ggaaggaagg	aaggaatgaa	61680
ggaaggaagg	aaggaaggaa	agaaggaagg	aaagattgag	gtaaggaaga	aaggaaggaa	61740
gggagggaga	aagggaaagaa	ggaagaaaga	aggaaggaag	gaaggaaaga	aggaaggaaa	61800
gatggaagga	aggaaggaag	gaatgaagga	aggaaggaag	gaaggaaaga	aggaaggaat	61860
gaaggaaagga	aggaaggaag	gaaagatgga	aggaaggaag	gagggagcaa	ggaaggaggg	61920
agggaggaag	ggagggaggg	aagaaggagg	gagggaggaa	ggagggaggg	aggaagggaag	61980
aatatttttt	aaaaagaaac	tctttgatga	ttcaacatca	tggagaagtt	attattatat	62040
tgagcttgac	ctggatctgt	aactcttgac	atltagctct	tgaataataca	tctgttttct	62100
tggctgatag	tcttaaaaga	cagaatccaa	cataataaag	ctgcacatgt	atccccatat	62160
ctaaaatgaa	ttttgaaatt	ataaaaataa	atacaatata	tgttcacaaa	tattaaaaaa	62220
atgtgcagag	atatgggaat	acatatgtgt	cttatgtgtg	cttattttcta	tgaagtcggt	62280
agatgtcaaa	gctctttctg	catgcacatc	tgcttagacg	ttttgtctga	aattctgttc	62340
acattgacat	acgtgcatta	aactaatatg	agaatgtatg	cattaaacta	gcttgaatga	62400
tgtccacatt	ctcagttttt	tcatcataac	tttttctttc	cagagtgcac	gatgccgcag	62460
aaaagaaaag	ctaaactgtga	aactcaaaga	aggtotttgt	tagggcctga	gtattttttt	62520
tttaattctc	atgactttca	cattcttagc	ttcagtaatc	cttgaagaga	acagagtcaa	62580
acctgtgatt	tttaattagga	tctttcagtg	tttgtggcct	tgaggagcga	gagcaagaaa	62640
gacattctat	gggtgtgctg	gaaagagagg	accaagggtg	ttttctttgt	gactagagat	62700
tatatgtgtg	aagagcagat	gacaaccttg	agattcgatt	tgtgatatga	aaccaaatgc	62760
actaaaagga	aagcgtacca	gtttggatta	gatatactaa	atcagtatgt	catgtaaatc	62820
agtgactatt	aaaagagaat	atataggaga	aaaagggtatt	agaagttgca	tttttttaacc	62880
aattgcaacg	tttgcaaatg	ggcttgtcac	ataccatcat	gtaagtacta	gtgaagtcta	62940
aggtataaag	aagtattttt	catttttaag	gatcattgca	aagatagaga	atltgacaat	63000
aatgaactg	actggccttc	tacaagtatt	tacttacaac	cacaattgta	atltttccat	63060
ggccaagcta	tgaattatta	aaagaatatt	gacattttgc	tatatatttta	aattatcttt	63120
aaatcatttc	aaatagtga	atagtattat	actaatataa	ctttattaaag	tacattttca	63180
agctcagcta	agttttttta	tcaatgttta	tttcagttag	gggatagttt	tgtttttttt	63240
cttctttgaa	ataatatctt	tgaaccaaat	gtccacattt	ctttacagat	aaaagagttc	63300
tgtttatgaa	ataagaataa	acagcctgta	atlttaatat	caccaaaact	atlttaattct	63360
ctttttggat	atgtacttca	acttaaaacta	cataatatagc	tatatagata	tagatatgtc	63420
tatatctaga	tgtagacaca	tccccaaaaa	cctatagaaa	tgtattttct	aaatgtgtaa	63480
agtatagtga	aactgccaga	ccacaagcga	tgttttgtat	ttatagatct	acttgcaatt	63540
tggggtgtgc	caaatcccca	caccttggtg	tcaatattag	tttaacaaaa	ctaaaaggaa	63600
atltaaaata	ttaagatcat	cctctttagg	atctacttgg	agcgctgaaa	aaggtgtgaa	63660
aatactgac	tcggacttct	tgtgagtttg	atgtgtttta	catttghtaat	gctatttctc	63720
tcactgcatt	catgcgtgtg	cttgaagctt	gctgccttgt	aggagacaat	gcaacaactc	63780
ttaccaatgc	cacttttata	agattttgcc	tgcaatgtgt	caatgtgact	cttcagtagc	63840
ctacttaagg	gagtgcaggc	aagatgacct	acctgagtc	atccaaagaa	tatttctcta	63900
gttctgttg	gtcctgtgaa	ttcctaattg	ctttataata	aatggccacc	taaaaattaa	63960
tgttaactgg	ttgattgtgt	tgcatgacat	taagtatacc	taactgccgc	aacagggtgc	64020
agctcgcac	atagcaataa	cgttacggga	gagaggggca	cgtcctccct	cagcgccac	64080

aacagggatc	attaaccaag	gctgacacaa	ttctccagt	agaacagatg	taaccttgga	64140
atctttctcc	acacagtgtc	ccaggaagcg	attcgacatg	agtctgtact	ggcatagga	64200
aataaaactg	gcattggctg	cacatagatc	ctattttctt	tgccctctta	ccaagagtta	64260
ttttaaaaat	gaaactattg	tcggccaggc	gtggtgactc	acgcoctgt	ccccagcact	64320
ttgagaggcc	aaggcaggca	gatcacgagg	tcaggagatc	gagaccatcc	tggtctaacac	64380
agtgaacccc	cgtctctagt	aaaaatacaa	aaaaattagc	tgggcgtggt	ggcgggcacc	64440
tgtagtccca	gctacttggg	aggctgaggc	aggagaatgg	cgtgaacctg	ggaggtggag	64500
gttgcaatga	gcagatatca	tgccactcta	ttccagcttg	ggcgaaaaag	tgagactgtg	64560
tctcaaaaaa	aaaaaaactt	agaaactatt	gtcaaccaat	gagtgtaaag	gtgttatttt	64620
tgcttttttt	taagtttatg	ttttgaaatc	ctctctcaat	tgacggaaaa	aaaaataagat	64680
acagaaatata	aagagacagc	ataaatgata	gaaaacaaaa	taagcatgtt	catgagccag	64740
aaatcaagaa	gaaaaggcaa	tgataaaaaa	ggctgaagcc	ggaggcatcc	caggcccttg	64800
gtctggaagc	aggtactcac	agtggcattc	ttgggcactc	tggggcaccg	aggtctgtta	64860
gtaccactca	taacgtgact	ggccagcaaa	atttactgct	taacataaac	aatgtctttg	64920
ccccacccct	gcaaaataat	gctgaaataa	aaggtgttac	tgatcattgt	ctgggactaa	64980
ggtttagtga	ggtttactta	cagaaaaata	aaaagcaaac	agtcatacac	ccaagatttg	65040
gccaaagctg	ctgttggatt	tgacatatcc	tgcaatcaca	gggaaaaggt	tggtactata	65100
ctgagatgtg	gctcaggatg	ggagaccaca	gggcaggcag	gaaggaggca	aaagcaatct	65160
agaagacctg	tgtgactagt	gagaaaaagc	agagtaagag	aaaggaagaa	agagagaaga	65220
tggtgtccct	cttgttaaaa	taaatgtact	aaaacaaata	gaacaatatg	aaaagaccct	65280
acataggtta	agtatcttcc	aaaaaagaca	cagcattcat	gtgagcttca	tattgctact	65340
atagcaaat	gccacaaact	tagtagctta	acacaacatg	catttcattac	tttacagcag	65400
cttagacatc	caacatgggg	cccactgggc	tcaaatccag	gtgttggcag	agctgctttt	65460
ctttctgaga	ctctagagaa	aaatygtttg	ctttggtttt	tcagctcttg	aggtgctgct	65520
gttccttggc	tcttggctgc	ctctccatct	tcaagtgc	tcactgcagc	ctgcgcttca	65580
tccatcacgt	tggtctctgc	gctcacattt	ctactctct	cttgcaagga	cccttgggat	65640
tctattggac	tcacctggat	aatccaagat	aatcttcatc	ttaaaatcct	taatttagtc	65700
acacacagaa	aatcacctct	gttatgtaaa	gtaacatatt	tccatatcct	ggaggaaggg	65760
ggcattgatg	tgtctaccac	agcatccaca	aagggaacaa	aaaatattgt	ccaataaatg	65820
catatagaaa	taaatttgtg	gattcaggaa	aaagtggagt	catggaacaa	aaatatttta	65880
tagatagaag	agtttccaga	ctagacatcc	aggaaagaat	tcacccagga	tgcagcttag	65940
agaaaaatgcc	aaggaaatag	cgaagggaat	gcatgcatgt	acatgcgcgc	gcgcgcacac	66000
acacacacac	acacacacac	acacacacca	catagaaaag	aacaattcac	agataggagg	66060
ggtagattta	gagactccga	catataacca	tgaggtttac	aataaaaaag	ataaaggaga	66120
aaacaatactg	aaaaaaagta	ggaattttcc	acaattgaag	aaaaacatga	gccttcagat	66180
aggatgtgcg	tgtctcgtag	tgagaaaaat	aaatgaaaaa	aaatccgcac	agaggcacaa	66240
agtaatgaaa	ttgtagcagc	tcaaaatcca	tattctcttt	gggtttgaca	ggaggaataa	66300
attcaagaga	tgtattttac	aatagggtta	ctacagttta	taatatattt	tatactcttg	66360
aaagtgtgctg	agtagatttt	aaatgctctt	actataaaaa	tgataagtat	attaaataat	66420
gcataatatta	attagcttga	tgtagccatt	ctacagtata	taagtatttc	aaaacaacct	66480
tttgtacatg	ataaatacat	agtcaaaaata	atttaagaat	ttttaaatga	taaaaaatac	66540
attttaaaaa	ttggaaggcc	atattcttaa	aaactaaaca	tgagaaaagt	attaccaaca	66600
aaaaataaca	attaggccctc	tctagaaacc	taactcttaa	aaacagacat	tacagaatac	66660
tttcaaaatg	ctttggcaaa	aacaaaaaac	caaataacag	ggaaagagct	ctatgaatta	66720
aaaattttgc	actcagataa	agcattaatc	aagatgaaat	acaaaatgaa	gtacatggtt	66780
tcttacagct	cagttttcct	atctggatta	actttagttt	tgcaatagta	tttcttgcag	66840
taacgttcta	ttatttgggt	gtccaaacat	gcatatttat	ccagcaacta	attcaaggta	66900
tgagagggaat	gaagatactg	atgacagaaa	ttagagcagg	gaatcagaag	agagaagtg	66960
atgttccctt	gcatggcact	cttcagttga	caaaaggaca	tgtgatccca	agatttctct	67020
tgactgagac	tacatttaca	caaatthaag	agttgtgtta	cattaaataa	ccaaagtcac	67080
agccttagga	acttccattg	caaaaaagga	gtttaacatg	gagctgttgt	taaaagggag	67140
gtcagatggg	cgaaaaatta	attggcaaac	aagttgattg	cataagctgg	aagtactgtt	67200
gaaaactttg	tggttcttta	ttaaacgaaa	acacttcgga	tattgggaag	acagcaacag	67260
tgaaattttc	ttggaatcag	gaataagggc	ctttatatat	tctactcact	gtctacatgg	67320
tcattgatctc	aagttattgt	ttctcataca	tatcagagat	gtgctaaaat	ttgaaatagg	67380
ttttcaatca	ggagccaatt	ctaatacacac	atcagtggca	acttggagcc	ttgagtttgg	67440
aaagattttt	tttagtgtgt	taggttggtc	ttgtgtgtgt	ataaagcaat	atctgagact	67500
gggttaattg	taaaaagaga	gttttaattg	gtcagcagtt	cttcaggccg	tacaggtagc	67560
atagcaccga	catctgctcc	tggtggagtct	tgaaaaagaa	ggcaaaaggt	cagcaggtgt	67620
ctcacatggc	aggagcagga	gcaagagtag	ggagaagtgc	cacactttta	aacaatcaga	67680
tcccttaaga	actcactcac	tattgcaagg	acagcaccaa	ggggatgatg	ttacaacact	67740

catgagaaat	ctatcaccat	ggtctaata	cctcccacca	agtctcacct	ccaacattgg	67800
ggattacatg	tcagtgtag	actcaaagaa	gacacatatc	tgaactatat	catgtacaat	67860
tccaagacaa	agtctactaa	agtctgagat	caattaaattg	tagatatctg	tcattaggta	67920
actgaaatag	aaggaatggc	cacagaggto	ttgtagggtt	gtggcttggt	cagggtgtctg	67980
ctgataaatg	tttaacaata	ggttttcata	gggtggtcagg	gaaggcctta	ttttagcat	68040
ttgccaatct	ccctgatata	aatattccaa	ccatggatga	tttactcaca	agatggaact	68100
gaattctgac	cacagctgtc	tcttgctagc	tggtgtgagt	cagctccaga	acctcaggag	68160
tctggtggct	taatgctgaa	acttagtatt	taaagttgtt	atcttttgtt	cacctgtacg	68220
gaagtaaagt	atgggtgtga	gaaaagtttc	agttagttta	gttagtattt	ataagaaaaga	68280
gtttaaaaga	aaagaatgga	aatggctcgt	gggtcgcata	attgagacta	tgaagagtgg	68340
ttgaataaat	tggtgtgttt	tatcctggag	aggaggggagc	tggagagtga	tttaataatc	68400
ttcatattta	tggaggtttt	ttttttttga	cagaagatgg	ttcccagttta	ctttccatcc	68460
ctaagtttca	aatcaagata	aatctagcct	taagtattcc	ccttttctat	ccaataatta	68520
tcgcttttaa	aatatttttag	aaaaaaatgt	actgtataaa	gaatgctttc	ttagattggg	68580
tatcagttga	tccagtttct	acactttaaa	tctgtctggc	cttttact	atatgtaaga	68640
ttaggcaata	aagataggga	aggttcaagg	gaaaacttct	gaggccaagt	aatccaatag	68700
gaatccaata	gcaatatggg	aaggcaggaa	ctattttacc	caataactct	agctagaaat	68760
ttcaacaatc	accaaccatt	cctgtcctcc	cacatgtcat	tcacacact	gcttcatgaa	68820
ttctaactcc	agaagtccct	ttgaattttt	cccagcctgt	cctgtcaaat	cgtgtgtcta	68880
gtgcaagcta	ctctttataa	tattgcaatc	tgttcttatt	ccacttttaa	taaacagagt	68940
agagaattgt	tgaccagatg	aataaaaagt	tggccaccaaa	tttgacatta	tccccctaac	69000
taacattatc	ttaaagtatg	cttctcagga	gggcatttta	agtcatttga	agttctaaat	69060
gcaagactga	atatcttctg	cacctgtttt	cattggttca	atatgaaaga	ttatactatt	69120
tccaccatgt	taccaaatat	gattggcatt	gtgttaaagt	ttcttaggtt	taactactct	69180
gaatcctttg	cctaaaatgt	agattttttt	ttttgtata	aaaaagacat	tagagtctat	69240
gccaaaattt	gcaattcttc	tataaaaatta	ttttttgtct	aaatcttaaa	aataacttaa	69300
aatcactaaa	tgttcaggga	aatctctagt	gagaaaaata	aaatttatag	ctcacagact	69360
agattgaaca	gacaaccatc	tgttcaactt	aaaacatttg	tctcaaataa	cataataact	69420
cacaaatgat	gatggattag	gtaacttcat	ttcatattta	ttaaattatt	aacccaactc	69480
tttttaaaag	catcaatttg	catgaagctt	atgagtga	gatatacaaa	aaaaacttaa	69540
tatttgatga	ctctctaata	tgtgctaggt	ttgttatatg	tatgacattt	taacaaatat	69600
ttagtcaga	tactatctct	tttttcagat	aaagaaattg	agacttaagg	ttaataattt	69660
acttacagtc	acaaaaatga	agtacgctta	tacaatctga	atctactgat	ggcttcaaaa	69720
tccttttcta	ogatgaaatg	tagcttcatg	taggggagta	taacatggaa	ataaggcagg	69780
tagtttttcc	tttccaaatt	tttagtatag	atgaagaaaa	agacattctc	aattatcatc	69840
actgtttaaa	aacaatgtct	ttgatttaat	atatggtaaa	tattaaatat	gtggctgtga	69900
aaataagtat	attttttagag	ttcaaagggg	aggaaaaatat	cctttgttaag	ctggggcatt	69960
ttcaatatc	aattttattt	attttttata	atgaaatagt	cagcactcca	accatccatc	70020
tatccacca	ttcatccatt	tatctacctg	tccatctatt	catccatcca	aacatctatg	70080
tatccatcca	ttcacacatc	cattatttca	tcctgcattc	atacatccat	aactactctc	70140
taagaaaagaa	gaaaaattta	ttcaggcttc	tctgtagatg	gaagtgtttt	gatactaaaa	70200
tgtggatgtt	tgtaatttta	ttattgccat	ttcctttcca	aatgtatgcc	aatcatacat	70260
ttttctaaga	cttatacatt	acctctgaaa	tgttctctct	tggaaatgat	gcacgtcact	70320
tctactgaca	ttttatttgt	caaagcacat	cacatagtca	tgcctgagcc	tatcctcagg	70380
aagacaaggc	catttgacca	ggaaacacag	gagatccaga	aatattgggtg	aacactggta	70440
acgtccatca	ggctgaaaaa	atgtggtata	aatattaata	tttaagacac	aaagaggat	70500
ttaccatgat	aaaacatatt	taagacaaag	gtttttacta	agagcaaaaa	cataaaactt	70560
gtgttagtca	gttctgtcta	ctctaacaga	ataccataga	ctgggtggct	ttacacgacg	70620
acattttatt	ctcagccagt	tctgatggct	ggaagtctga	gatcagggtg	acagcttctc	70680
ttgaggacca	tccttctggc	tgacactgc	tgacttctcc	ttgtatgctt	aagtggcaga	70740
tagcagcgac	ttcctggagt	gacaagggtc	tcaatcccat	tcgcgagggc	tccgcctcat	70800
gagctaatac	ccaaaagtct	ccacctccta	ataccaacac	cttgaggatt	agaatttcaa	70860
aatatgaact	ttggggcaca	caaatattta	gtocataaca	gccttggagg	taaaattcta	70920
agtagcagaa	atctttcatc	tcttctaaaa	ctaactag	tggaaaacat	aatctttcat	70980
tcaacgctta	gagtctgtct	tggagagcag	gaacagatgt	attatattac	atgtgcacac	71040
atacacatat	acacacaaac	atgtgtttgt	atgtgtgtgc	ttatgtggat	atctatatgt	71100
catatataca	ctgcactctag	cacaataggt	aattgctatg	tcagggcaca	tttattatag	71160
ttttctctct	aactcataag	ggagagtttt	ttaatcagct	ggtaatatca	tatatttcag	71220
ttttttcttg	caaataagag	tggaaatagaa	agattttcgc	tgttgagaat	tgtgccttca	71280
agtgttaaca	atattcacaa	gtgtttaaag	aaataattag	gttcatttca	ttctatcatg	71340
cttgactat	atttataaca	tgacctttgt	tctatttata	tatacaagta	atatattcta	71400

ataggaattt	ttttgtgaag	aagaaacact	tgataatatc	ccaattccct	aacagaatta	71460
tacaatagtt	cctcaatgac	cttaccttgg	gaaaaaatta	aatgaaagac	acctgcatta	71520
gagatcaatt	tcctgatgct	cagtttatat	tcatcatgct	aagctatcca	aggacacctt	71580
tcttgccta	tcatttgatg	gaaacgattt	ccatgtgttt	tatgattcaa	ctccattctg	71640
cottagacaa	tacagtcact	tcaaacagtt	caactctaca	gttggtctggc	tgactttgcc	71700
ttgtaaaaa	tgcttttttt	ttttttgggt	cagagtattc	tatatatttac	ctaagaagtg	71760
gaaaacaaat	tttgggcaag	aatagtggca	tacacattac	agagtgattt	cacatttcaa	71820
tattattttac	aaatatgaca	tcaaacatta	gtggttttat	ttcattttta	tatcagaggc	71880
attctaaata	tgatgtgaca	aagttaactc	cgaatttcct	ggtttttgtc	agtttaaccc	71940
agatctttgc	tatgcagttc	tgaacactgc	catgctttca	acttttcaac	aagaaacagg	72000
tggttaatac	acacatgtaa	ctctacaact	tgatcttagc	atttcatgta	atactttggg	72060
aatgcctaaa	tattatgttt	gtagttgtcc	ggagcatctt	attttcactt	tttcaattct	72120
atcaaaaggt	acatttaatt	aaaatttaacc	aaaactgaat	aagaataaaa	cctcacttcc	72180
ctaattttatc	ctcttataga	aaaatatttt	cagtgtttat	atttatgtga	taccagccgc	72240
atctgggttc	agtaagcacg	tgtttcgagc	atactttgag	atgcattgtt	ccttggcgca	72300
cctcagtttg	tactggaatc	attaatgcat	aatctgaagt	tgtgattatt	gagatttttc	72360
actcaataat	gggattcaaa	tgctgccctg	tttgatctgc	attaactgtt	tggaattcca	72420
gaattattgac	tggcacttca	gaagtcacca	agatattctg	tagccttcag	agggtaaatt	72480
tcagtcacgt	ggggatgatt	tttttggcaa	agagcatttt	gcactctgag	tcctcacaca	72540
cattctcaca	agtttttaag	cccaggacag	tctccattcc	accacttgta	ggtaaactgg	72600
tgatgatttt	aaaggtaatg	tcagcctgag	aaacagtga	atatgctgaa	gatattggaca	72660
caggcacgga	gtctgcgtca	tcttagtccc	taactctcac	cttagtgaat	tgttccaaca	72720
atgacataat	gaatttatatt	taacttttcc	tgctgctcat	gttttgatat	ttgtcatcaa	72780
tcggagctgg	tgagagtttt	aataaagtgc	ttagcatcac	ctgagggtaa	atttctatat	72840
cctttggatt	ccttgatcct	cctctgcact	ttatgtaatg	agtgaacctt	ctatctagtt	72900
caaatagaga	atgtagagaa	gtaaagtgc	tcagtcaaga	gggcatcact	gatgtcaaaa	72960
tgccagggcca	acatgaaaga	agggcagaaa	ggttccggac	acaaaaggga	gggactgcc	73020
tggaactggct	agtgtcgggg	tttcagccag	aaccttccct	tagaatttgg	aaactatgga	73080
gaaaacacag	tgatgcagat	caatatatac	cagttttgtt	acagatgttg	aagaagggtg	73140
ctcactactg	tgatcataca	tttgggcttc	ccaacttttt	ctcttgaatt	aaacatatta	73200
ctaaattggc	aggctttttt	cctaacaact	gtgaaatgct	agtctttaga	aattttcagc	73260
atccagcagc	aaataacatg	catgtccaac	ggcgaggatg	gctccaaaaa	agagagaacc	73320
ccaaatgaaa	atgagtcagg	cattcttttt	tctttaagca	gagcaagatc	attcttatca	73380
gaggaaattg	tgagcagggt	acaagaagct	gcacatgcct	aggagggtgag	agttctgtct	73440
actccaggga	aaaatgagaa	gtgttgaaca	agcagcactc	ccagaaaata	acacattgta	73500
gtagacgccc	gttacggttt	ctccatgtat	taaattgggt	tataacattg	tctttttgtt	73560
tgtctctttt	gagtttataa	gttcccatca	atattttatt	tcctttactt	attttgggtt	73620
taatttgtgc	ttttttctat	ctttaagata	aaagcataga	tgataaactga	tttttagactt	73680
tgcttatttt	ctaatatagg	catttaaaga	tgaaagtttc	ctcttacaac	tgatttagct	73740
gtattcttca	aattttacct	ttcaaaattt	ttaggggttt	atttttatta	ttgttcaatt	73800
tgaaatattt	cctaattttc	cttgtcagtt	ccacttttac	atacagatta	tttaaaatta	73860
tggtgtctcc	caaataattg	ggaattttta	cagatatatt	atactattgt	ttgtgatttc	73920
taatagaata	ccatttatagt	cagagaaaaa	tgatcatgtat	tatttcaatc	cttttaaat	73980
tatcacaact	tgctttgtgg	ttgagatggt	gatctacctg	agggaatgtt	ttatatgcat	74040
agaacagata	aagtgtattc	ttcaattggt	gaatataatg	ttctatatgt	gtcatttaga	74100
ctaattattga	tattgttgaa	gtcttctagg	tccttattaa	ttttttatct	attagttata	74160
tttaattgctg	aaaaagcatg	tttaattgtct	gtgattggag	tctttttctgt	aattctcctt	74220
cagttttttt	aattgttaca	tttatgtagt	ttaaagcttt	gtgattatat	ttatccacat	74280
atttaattgg	aaatttttgt	tgttttaaaa	atgtttaccc	tttctggggc	tttttcttcc	74340
ttccaacctg	gtattttttc	tcttcagcct	gaagaagtta	cctcagcatt	ttttatagaa	74400
caggtctgct	ccagacgatt	cagatttttc	tggtgaacat	ataaaaaatat	acagaaatgc	74460
ttttaaggca	ttgtctcagg	atatttgcta	tcattgtttac	agtcttttat	cttctactat	74520
ctttcaactt	tttttttttt	tttttttttt	gagacggagt	ctcactctgt	cgcccaggct	74580
ggagtgcagt	ggcgcgatct	cggctcactg	caagctccgc	ctcccggtt	catgccattc	74640
tcctgcctca	gcctcccgag	tagctgggac	tacagggcgc	cgccacccgt	cccggtcaat	74700
tttttttttt	tttttttttt	tatgttttagt	ggagacgggg	tttcacagtg	ttagccagga	74760
tagtctcaat	ctcctgacct	tgtgatctgc	cgcctttggc	ctcccaaagt	gctgggatta	74820
caggcgtgag	ccactgcacc	cggccatctt	tcaacaattt	aaaaacattt	tatggtcttc	74880
tagtttcatt	catcttatga	gcacaatgcg	tgttttttat	ctcctccagc	ctctttcaag	74940
attttctatt	catgttttgt	tttcatcacc	gatttgatga	tcgtttgcct	agttcttggt	75000
atatttgaat	gtattttactg	tttcttatat	atgtgagttg	atatatttca	tctaatttag	75060

aaaattgtca	cccatgttct	cttcaaaaat	gtcttctacc	tcattatctg	tcctcttttt	75120
acaagaccct	taacgtaaatt	atatttgatg	gcataatatt	ttctacacat	ctcagatggc	75180
tctccttttc	gttttttatt	tatttttatt	tatttttatt	ttgagaggag	tctcactctg	75240
tcacccaggc	tggagtgacg	tgggtgcaatc	tcgggtcact	gaaacctcta	cctcctgggt	75300
tcaagtgatt	ctgctgcctc	agcctcccga	gtagctggga	ctacaggtgc	ctgccaccat	75360
gcccagctaa	ttttgtattt	taatagagat	gggttttcac	catgttggtc	aggctggctc	75420
cgaactcctg	acctcagatg	atacacccac	cttggcctgc	caaagtactg	ggattacagg	75480
catggctctc	ctttccttct	tctcttctgt	atttttctcc	ttcttctctc	ccatctctcc	75540
ctcctcctt	tttttttct	tatatattcaa	tgtgggtgat	ttctattgat	ctatcttctt	75600
gtttactgat	tcttttcttt	ggtatgccag	tttactaata	gacctgtcca	attaattccc	75660
catctcaata	ttttattttt	tttccctatt	cttgacattt	ggcttttttag	ttgtgtttat	75720
ttctctgcta	aaattcttca	ttttttttct	gaaaaattgt	ctaaaaatag	tgtagattta	75780
tctgacatat	ctattatatt	taataataaaa	ctcctctctg	ataatttcaa	catctatatt	75840
tggagatgoc	tctatttctc	tttcttccct	tgcccatatg	gcatagttat	gtttatacat	75900
cctataatgt	atatatata	agcttctcca	tgtcttgtaa	ttttttacga	tcccttccat	75960
ttacacagggc	atgagatcta	agcaagggag	agcctcggat	ttctctttgg	gctttataaac	76020
ccagctgctg	actccctgag	tcggggcttc	cagtgaattc	tctttgctgt	ctgggtgctc	76080
accagcctt	ttcacaaact	agattttctac	gctttgctgc	tcagctctga	gcctttgagg	76140
cttgagtggg	ctcagtgaa	tccccgggag	ccttgagggt	gtttcccaact	gctgtcctac	76200
tctaccacgc	tttccgcaac	agctttttgg	gtgacagggt	tttctcatgg	ctctcctgag	76260
ctgctcccg	cctcaagtca	gctatataag	cctccatttt	tcagatcccc	tactgtgccc	76320
ccaaacttgg	gtgcattatg	tccatgtact	aaagtaaagg	ccccctggga	cagagagggtg	76380
ggtgggtgca	gagctgcctc	tgatggagct	tcctcaggga	cctgagtcac	cacacttcat	76440
aagcttttta	aaatgtgttg	aaaattggac	ttgtttctca	catccctttg	ctggcaaatc	76500
cctcctcctc	ttcccatcac	tccacctagc	atcagggaag	cacttccatt	ctctctcata	76560
ggaagagatc	atgactttct	agattttctt	tcagtttagc	tttttttgca	tctccagctc	76620
ttcattggct	cttataaaact	gtgttttgtt	attgttgttt	tcaagttatc	catcttgttc	76680
tcattttag	ttctcttatg	gctttctaat	tcctaaccag	aagtggaaac	tgctagctat	76740
tgtttataac	atgggtccatt	tgcagtttat	aggggtggta	tatgaatcaa	gtgtggaaag	76800
gaccagaaaa	aaattgtaat	ggcagatggt	atgttgtctt	aagccttgct	ttgcttttgc	76860
cactgaggct	gtgttgtttt	taattcttact	gtgcatggca	ccttctcatc	acactagtgt	76920
tttgcccttc	aggggtatca	tccatataaa	taaactgctg	gaaggccagg	cgtgggtggct	76980
aacacctgta	atctcagcat	tttgggagat	gaggcagggtg	gatcacttga	ggtcaggact	77040
tcgagatcag	cctggccaac	atgggtgaaac	tcgtctctac	taaaaataca	aaaatgagcc	77100
aggtctgggtg	gcacacgcct	gtagtcccag	ctactcagaa	ggctgagget	ggagaatcac	77160
ttgaacctgg	gaggtggagg	ttgogggtcag	ccaacgttgc	tccactgcac	ttcagcctgg	77220
gcagcagagc	aaaactccat	ctcaaaaaat	aaaagtgaat	aaaaaataaa	aataaaactgc	77280
tggaagaagt	ttatgaggta	tcctaagaca	tgtacaaata	ctagtgtaa	acccaactgg	77340
gaatctttat	gtcttatttc	gtccacagct	ccccacccac	accatggatt	taagaaagt	77400
ttttaaaaca	ttttagatgc	aaagctatctg	gttctaaaa	attttagat	caagctatct	77460
ggttttcaag	gaatttatat	taagattgat	tttcaaatcc	tgtgtgctct	agtctaggag	77520
gataattatt	tyttaatgta	gatatttcat	agagacttca	ggcaaacaca	tgaatgagga	77580
gcaaggaaag	acccaaagaa	agcatgtaag	tagatatcct	aaatcaaatt	gttattttgac	77640
caaagaattt	ggtgatatga	taaaatccaa	tttctaaata	atgagttttt	ctgttaaagc	77700
tctcaatatg	aattaggaaa	gccttcagct	acacgtaaaa	gaaaaatctaa	ctcatagagt	77760
tttaaacaaa	tgatgagtct	tttctttttc	cttaaatctc	ctggaatcct	gttgacattg	77820
gattagcaac	aatggtgtca	aggactgaaa	gctagtgatt	gcttttgcct	tagacttcat	77880
gcctacattc	aaaacaagaa	gaggttagga	tacagcgtca	ttcattctgc	atcatcagga	77940
aagcaaaaag	ccttccatga	agcctcttgg	cagtcatata	cttaacacct	ttttggcaga	78000
aactgggtca	tgtgtccgtg	tctaggtgtg	agagagggat	ggaaggtggg	aacatggagt	78060
tgttaggcca	attgtatttc	atctcttggg	aatgggggga	ttgatattgg	gtagacaaaa	78120
atcagttggt	atcactctac	ctaagtgttc	ctctgaaata	tattacttga	ttctggaact	78180
ttatttgaaa	agttatttga	tattcaatag	ttataatggt	aatccaatac	aaaatatata	78240
tggtgctttt	aaaatgataa	atagcaattt	aatatctaga	gtaatctcat	ctacatttga	78300
cagtaatat	aactattcta	ttctcataaa	tcttgtatca	tttttttgtt	aaattgtcat	78360
taattttaagt	aaggaaaagt	agagtttatt	tcccagtcac	aaatagaagt	aaccggttgt	78420
tgtaaaatta	tcaatgtttc	tttcttgaac	cactgcacct	tgaattttga	ttaaagagtc	78480
gaagtattgt	taacatccgt	ggaactttaga	aaaaatacca	gtgctaacca	catcacatct	78540
aaaatataag	cacatttcaa	atttatgtga	atttttaata	atctgttttc	tttggttgt	78600
agaaaaatta	ccaaatagca	tattgtatta	agaatgaaat	gttctgtaat	caagatcaaa	78660
ttacaatctt	ataataataa	cagataattg	tatacatgca	atattaaagt	atatgagcat	78720



attcttttaga	caaaatatcc	aaagacatca	atgttttcat	tattaaagag	ttactttcca	78780
gttcactgaa	taaatgaaaa	tattaaagaat	ttgggttatat	taagaataaaa	ctctaagatg	78840
tttcttttac	ttcagtatga	aaagcatgca	aaaatgataa	ctacattttg	ttatgatggt	78900
ttgaaaaatc	tggcataatt	taaaagcaac	tttccatctg	gctggtgcag	taagcaagtg	78960
gaattagaag	agtttggtctc	tgtttcttga	tgtgaatgat	tgtgattagt	ctgatttctt	79020
tggtttgctt	ttattactgt	tttcacttct	tttcttggtg	tatgtgtgtg	gatatgcata	79080
caggatgaca	tttctatcac	attatggctt	aacaaataaa	acagaaaata	tgtttacttt	79140
ggatgtgtgt	gtatttttgg	agacggatatt	tgttgcctac	tcaacagtg	tctttttact	79200
tcttggcagc	ctccttggtg	taaccacgca	tttacggaag	atgacgttac	cagcagccct	79260
ggggttaagtc	tcagtcaccc	gaagccagtc	ttggtaatcc	cagcagacta	gcaagggaac	79320
acaacctagt	tctgtcaagt	ggaaagtga	gagagctgtg	ctgggagggt	tctcagaaaa	79380
ggtttctttt	ggcaaaaagaa	agagaggggg	gaacagttct	ttttcttctg	ccggacaagg	79440
tcatgtgtgt	acgttaagaa	tgggaagttct	catgcccact	gagcatgggg	agagtctctg	79500
ctaggtgtac	ctgggtgagaa	cagacaaccc	ttgagccacc	ttaccctagg	aactcctggt	79560
gataagacag	aatacatttt	ttctcactgt	taaactcttt	tttattattt	gtagctgggt	79620
gcctccaaac	caatataatt	agcogtatgt	aaatattctg	aaatctctgt	caattcaaat	79680
aacatttagg	gtttttttcc	cagtgtatcaa	aaatacagca	tctcaagcaa	tattgtgtgt	79740
gtgtgtgtgt	gtgtgtgtgt	gtgtgtgtga	tattattaca	taataattgg	aagacttttg	79800
atgtagatcc	tagtgggtgat	cttaaatggt	gagtatttga	cttttttttag	catttaataa	79860
ttgtcacaa	aaattttatat	ggagagatat	tttaattata	gaaactcatt	agtttgtcca	79920
taatgtaaaa	gagttaaata	atatagttct	gttgataatg	atattttgta	atagatattt	79980
ttaagctgtt	aggcatgaag	gaaaataatg	ttgtaagtaa	tcagttttat	ataggggtta	80040
tagatatatt	tattatttgg	tattgtcatg	tgttttcatg	cctcataata	gctgattaca	80100
aatagatttt	aatcaatcac	tctctttatt	ccaaagggtt	aagcagataa	agaatacaga	80160
gaatgcttgt	aaagcatttt	ctaccaatat	attaggacca	tttctggata	aggtccttga	80220
cttcttcttt	taatcgtttc	taaaataaca	tttccaaaaa	cctcaagcaa	aaccacttta	80280
aaaaatctga	attcatataa	ttttacctaa	tgtggaagat	tgggagagca	gtgactttta	80340
tataatattt	taatgtcaat	gtgtgtacat	ctgtaattaa	taaaattagt	acaattcaca	80400
acacatggtc	gttcttaagc	gacagagaca	tagaattgta	acagcacatt	aaaactcctc	80460
tctagtggcc	ccgagctgca	ccccataaac	tctttgatag	ctcttaacaa	atgcaaatata	80520
atgagcatta	acccaggcat	tcgaaggctc	agataaatca	ccatgctctt	caggaattca	80580
tatgaacaat	tcacagtgct	ataatttttc	ctaagttaag	actagactaa	tgttgagaga	80640
gacagaaaga	gaaacaaaaat	ataggtgcac	tctcctctgc	tcattatagg	gttctcttcg	80700
acottgagga	gactgctact	tggggacatg	tactgattca	ttttctgact	ccttttgtcc	80760
ttgattttga	catagccttg	gtgttgggaa	attcttccct	caaagcctca	atcgctcatg	80820
ctcaaattta	aacacattttc	cttttctgta	atcagtgagg	gaagagcagc	tggtcaccat	80880
cttctacaca	ctaaacctttt	gagccttttc	ttcagaataa	atgatcctac	ttcttagaat	80940
cataaaatta	gaaggggtttt	tgatgtcaga	ccatcctttt	tctgccttta	gacaagacta	81000
cccttccctc	ttaccatccc	atacagatga	gcaaaaccat	cctattttta	aaagagatca	81060
agagaaaacta	atcaaaagcc	tccaacagtc	acacattcct	ctgatgagca	acccttactt	81120
tcaggaaatc	atcatcatta	tcatcaaaata	gcattttcct	aacacccaca	tgtgcaaaat	81180
atgtttgtgt	attcagatcc	taaagagggc	cctagaggcc	cagaaaaggt	tggaaaaaga	81240
aaaaaaaaaa	acgtgcaaaag	gtacacaagc	atcactaggc	tctttgcaag	agagtctagt	81300
tcattgcaat	aataaattat	ttttaaaaga	cattccagaa	actccagggg	attttcattt	81360
tctatttagg	aaagcgatct	tataattaaa	gaatatatcc	attcattttc	ttttgtttaa	81420
ctattaggac	ttagtaatga	ctatcaaaact	caagtcta	gaccaaagat	tgagaacata	81480
aatatgcagt	gtattaaatc	tgaacaaaat	ccatactaag	gtttttaa	ttgcaacacg	81540
tataatactg	aaccagttgt	ctctgttaca	aagtgtgtat	gtattgaggg	gaattgtctg	81600
gaaaaatgtt	taaagctctc	ttctctccca	ttctggctcc	tcctttatgc	tcctggagaa	81660
cacagctcaa	aggctgctct	caaatgttca	aaactcccca	cccatggctg	gaatggctot	81720
atgtcttcca	gtttcacott	ccagtcagg	aattcccaat	ttctgagtat	tactgatttt	81780
tgctagggtt	taactaaata	aagcaattta	attgaaggta	agccttccct	taatttataa	81840
atatttgata	aagggaat	ttgccagtcg	aacatgctgc	ttgatgcggg	gggggggatt	81900
ttgacctggg	tggcactcca	gtctggactt	caacttacct	ttaattgctc	tgccagggga	81960
tgcaacacag	agtaaggctc	cacgtcatcc	attcactcct	tcctcttttt	gggtttctca	82020
gctcctctcc	attgctcacc	ttcacgagaa	gttccctgta	accgtgaagg	ttatcacaga	82080
gctcagaggg	taggagctga	cccagcacct	gcagatcctg	aaggggccct	aatctaaata	82140
caccattcac	cccaacaggt	gggaacaat	ttgatacaca	ggctcaaaag	tttgttttga	82200
ttactgttgc	tgtgaagttt	tgatttggcc	tcctaaagta	gaggagccaa	tttgatattt	82260
cattgcataa	ttcctgtact	atagtagtag	gtactatcta	gcaagacttt	actgcaaaaa	82320
taaaatatat	taatagaaca	tacaatggta	ttggcataca	gctcttttta	ttttttatct	82380



tttctaatag	atatgtttat	tctatttcct	aaacgtatcc	aagacttaaa	attttcaacta	82440
ccaatgaata	acatgcttaa	atatttttagg	gaaatggtaa	aatggtaaac	agctctatta	82500
gtattattga	atatctcata	tttgaaagga	aagaaaagtga	cacttgctgc	tttctcatatc	82560
gctatcacat	ttaaacctaa	ttacatttca	agacaagaat	atztatcatc	tttttttctg	82620
ttaaagaagt	tgagcctcct	ttacgataaa	ctgaactcaa	atgtaagcat	ctctaattct	82680
ggagtgcata	tgtgcacaat	tacataaata	atttggttta	agaatcaaca	aaactaaata	82740
aatgttaggt	ttctagaatt	tatccaaaca	tcatgatgat	tgactattta	ttgccgattt	82800
tcctatttga	atatgcagaa	gtttcaattt	taagatataa	tcgttatgtt	ctccttattt	82860
gatctcgtat	ttccacagag	ctactctgtc	atthaagtga	cctaaattct	tctccatagc	82920
cttttaagat	tttccgatgt	ttatctgttc	tctgtatcag	tctcgatcac	tgtcttcaact	82980
ctaaaagctg	cacttaaatag	cttacttacc	gaattagtaa	tttggaatt	accgtatagt	83040
tacatggaaa	tattacttcc	ctctgtgcca	cagactgggc	tttgattact	tcgatttcaa	83100
tgccccagca	gcattccagat	tttatgaata	ccatttccga	tgctttgctt	aaagttaggg	83160
gccaggcctg	agctcaagtc	tcaaggctga	tgttgtaaat	atcaggacag	gcgggttcaa	83220
aagcacaggc	gccagctgcc	ccacctccat	ggcacagccg	tcatggagcc	agccacaggt	83280
ctgtttcaga	ttgcaaaagga	gggtctctga	cttctgtttg	cttttgtgct	ctaggaagac	83340
attgtaggcc	tccataaaca	aaaataaaat	gaaatcgaat	tacaacttct	ttttaaccta	83400
tttaggatgt	aaaggttaaa	tatatatgta	tctatgaaga	aatgcatgtt	aaaaagtgg	83460
ttgaaaataag	acatgcggaa	gcaaaattca	tttttctttg	tatgggtaag	tagtatgttt	83520
ttctttgtat	gggtaagttag	tatggctaag	tagtatgttt	tctttgtatg	ggtaagtgtt	83580
atgacaatac	gccataatca	ttgtccttat	ctattaacta	aacctcaatg	agaaagataa	83640
attattacat	tgagagagata	tgagacttaa	gacttactca	aaaaatttta	aaggctattc	83700
cttaatgtaa	tacaacttcta	tcaggaaatag	tacagaattt	tatctgagat	atttgtatta	83760
atcaatattt	ttgattcata	tgaatcatta	tttgatcag	attttgatcc	aaataaatgt	83820
gaagtatatt	gaataattca	gaattaatta	gtttccattt	aatcagttca	atccttctga	83880
gttcatgatt	aggcaatttg	gaaaagacta	tataaaataa	tattctattt	gaactatgga	83940
agtattcaaa	agaaactggt	aagttcatga	gtataattgt	ccaagtgagc	tctgtaatct	84000
atgaatatat	aggaaagtct	tgtctttgag	aatacatggt	ttaagtgaat	tctattatct	84060
atattgtcat	tttgattgta	tttcaatttc	gttatgtgta	aacggaagca	atggatgtac	84120
tctttgaaaa	cagtttcaca	aattaagaac	tttaaacaaag	cttttatctt	ctatttaggt	84180
caattttgtct	aggtatttga	ggtaattttc	atcctcatct	ttttatttgt	tatcttgtcc	84240
aatcctatta	atgagcataa	atggctcccc	aagtgtaaat	atttgaagg	aaagattggg	84300
tcattgtaaa	ttcaggtagc	atccatagtt	gatgacagaa	tgacgggaag	aaactttcct	84360
cactaaatgt	caatcaaatt	catcagcatg	atacttttac	atgtctgcac	agctggaaaa	84420
aacagtcaat	tctgtttcat	tgttatgtaa	atztatgctg	atactcaatg	taatcaacaa	84480
ccgagtgtag	aaacatcctt	gaaacagatt	agatgattgg	attcttttta	ataataattc	84540
gaaactggaa	ccctaggttt	ttcttaggca	cttcttaaat	atztatggtt	atccagctta	84600
agtaacacag	aaattctgtg	gtcagctagt	ataatctttt	tcaagcagtg	tatgctattc	84660
ttataccac	accaacaaaa	ttgtatctaa	aacattggta	tctaacctag	gaaaaatctc	84720
aggtaagggt	ttacgattca	gtgccttagt	ccccttaagt	catttcggcc	acagccttaa	84780
aacctctctt	taaaaggatt	taccatcacc	ctttgtaaaa	actggatttg	agcaagatgc	84840
attttattaa	tcaaatccat	aaggcagtca	atacattact	tcttttgga	tgaaaaagga	84900
aaagagttat	ttcacaagca	gatgaagact	acatatgtaa	ctctacttcc	ctcggtaggc	84960
tccaatacag	gtttaatcct	gaattttccc	attttgaaag	gcttgccaag	tcagttctca	85020
gacctttcgt	gaaacatgtt	gctcagaagt	gacatcaaac	gtgatcctaa	tttttgacag	85080
attcttctat	acaatgacac	atacctggcc	ttcaaaatag	agtaaaactaa	aactcaactga	85140
agaagaacaa	aaaggtcact	cttaccattc	ttgaggattt	cgaaaatcaa	tgatcctaca	85200
gaaggagaag	aggccattta	tcattatagc	ctttttgatg	tcaaagaaga	aaaagaaaaa	85260
gaaacagaaa	aacatgcctt	taacatagca	gctgatatta	gtagctttag	aatttacaaa	85320
gcactttact	gttccgttat	ttcaattaac	actcataatg	ttgttgttca	gccagtctta	85380
acatttgtgt	tataattatt	tcctttgaaa	gctgaggaca	ggagggtcaa	ataacttgta	85440
tcattttccca	aagtccttgca	gctgataagt	gaggttctta	ggagtctatc	tccgaaagat	85500
tgagcatgaa	atctgttttt	tttttccatt	ccattgaagt	gtgtatttac	taactgggta	85560
tataattatg	aattgaagac	taacctttat	ttaaattagt	gcaatcagac	tatatccaga	85620
atatttctag	aaatgatatt	taagtatttg	atgagtaaaa	tgaagggtaa	aatatttttt	85680
gtttgttcac	ttaaatgcaa	acctcttgag	ttaaaatttt	ctcaataatt	atgtagttca	85740
aacaatattt	tccaagatta	tttgaactta	gatacaatta	ctgaatttag	gcatatatac	85800
attagggttt	cttaactcttg	gcattcttgg	cttttaagt	gaataattat	ttgttttagg	85860
aagctgtcct	gtgcattgtg	ggatttgggg	catcctgccc	tctacttacc	agatgctaac	85920
agtactccag	gttgtgacat	cctcaaatat	ctttagaaat	tgccaaatca	tacttaggtg	85980
agaatcaatg	gtctgtatta	atttaactgc	aaagacaagt	gctgtttcta	aatattacac	86040

tggacaattc	taaagggttta	ccataaaatta	gcagaggcat	taatctttta	atgatagtat	86100
tgtagtggat	atactgggtac	aggacatagc	tatctaagct	aattctatat	atacagtaga	86160
ccaaaactga	cacattccag	gcattcccttc	taattctcct	ttgactcctc	ttatctagag	86220
tgaaaatttt	cttagcttat	tttgcatctt	cccagcatat	cttcaactggg	agaaaattcg	86280
tctctcacat	ttttgcacat	cttgcaagtc	aggcactgac	cacctttttg	ttctggattc	86340
tcttttgagg	atgtttatat	agcaaacatc	cttggaatg	aagaatcatt	tctgcctgtg	86400
aaacagtggt	caaacatact	tactccccat	tatagaacat	tcaggttccc	taagttcagg	86460
ttcctcttcc	cctataaatt	cctttgcccc	ttgtaaaatta	ttgtttttaca	tgatcctctt	86520
tgtgttgcca	tgtgggaaca	aaatggtgat	actctggata	ctgctgttac	taaaacttttc	86580
ttgatctctg	tcccaggagt	cttctcattg	cttctacta	gaattcatgt	atccatatta	86640
ggatacattt	tttgttttgc	cagaaaggca	aaatccccag	tccttctcat	tactatgttt	86700
ttgttgtttg	tttgtttgtt	tgtttttatg	cagcctcggc	gctctgtcac	tcaggttgga	86760
gtgcaatctc	agctcactgc	aaacttctgcc	toctggattc	aagccattcc	tctgcctcag	86820
cctcccaagt	agctgggatt	acaggcgtgc	accaccacgc	caggctaatt	tttgtatttt	86880
tagtagagat	gggttttcac	catgttagac	aggctggtct	cgaactgctg	accttgaaaa	86940
tctgtccttt	gaaatgaatg	acaaattaca	tgctattttat	gccccaaaat	gctaacttcc	87000
acatatatca	tttttagtata	tocttataga	ttatctocaca	tgaaatttct	ttttgcaccc	87060
attgaaagat	tattcctgag	gatgtcatcc	actctatcac	ttattgtttc	ctctttatat	87120
aagcatattg	tctttttaaa	ggaaagctga	tttttacctt	aacaacagtt	tcatttcttt	87180
atagacatca	caaaagttgc	gagttactgc	atttgcattc	tgcttatctc	ttttaggaaa	87240
tgaaggcttc	cttttttaag	ccaacctcta	gcagctgcat	aatatcattg	atagtaatgt	87300
tgggtgttaa	ccttaccact	gattttattac	ttttctactt	ttaaataatt	gggtatctga	87360
ggaaatggaa	ataagatata	tattaaatat	agggtgtgtg	atatatatat	atatatatat	87420
tcaagtatgt	atagtttcat	gccaacctcc	atattctata	tataaatata	tattacatat	87480
tatatattat	atatgtatgc	gccaacatct	taaaaagttt	acatagaaag	acatcttgga	87540
ataaattgct	attattagt	gagtgggggg	aagaatcaca	tgagcaatct	aggtagcaat	87600
atatctaccc	tacatgaaga	agatcacatg	atctttggtg	tttctcattg	agtatccggg	87660
ttagtagata	acttattgac	acagggatcc	cagagatcct	caggggaggaa	agtgtattgtg	87720
accccaacac	aaagcttggg	aaagtgatgg	aatggactag	agtggcactt	gacaaactat	87780
tctgagagta	gcactccagg	agcattttgga	gaattcacaa	ggatggtaac	attaataaca	87840
agggtggctt	gatactagtt	ttcaacacaa	gctctggccg	tcctaagaaa	aggctttttc	87900
tggccaaaag	tgttgcttac	tctggaaatc	cctaagcact	ggccattctt	gcatgaagtc	87960
tttcttcggg	aaaacatgcg	taaatgattg	aatataaatt	caaagaattt	taattgttga	88020
gcccaggtat	agagtgactt	cactottctc	tggggactga	agtggaaaga	ccaagtgtcc	88080
ttcaggatgc	gtggactcca	tattccacaa	cattgtgtgg	cagcatcaga	agccactgcc	88140
ccagtcgggt	aaaaccaccc	tctccacctt	ctccactcta	tctttctata	aacaggatga	88200
ctgattccca	tacactgaat	tgaccttgga	attcatggac	attgtggttt	ttgtttgttt	88260
ttgttcgggt	ggttgttttag	ttgattgaga	gaagtgaggg	gaaatgataa	tatcagatgg	88320
ggagaataaa	ttgattaaaa	tttcttaaaa	ggggcgggt	atggtggctc	acacctgtaa	88380
tcccagcact	ttgggaggcc	gaggtgggtg	gatcacctga	gatcagaagt	tcaagaccag	88440
cctggctaac	atagtgaac	ccagtctcta	ctaaaaatac	aaaaaactag	cagggtctgg	88500
tggcggtg	ctgtaatccc	agctgctcgg	gaggctgagg	caagagaatt	gcttgaacct	88560
gggaagcaga	ggttgcagtg	agctgaggtc	acgccattgc	actccagcct	gggcaacaag	88620
agtgaactc	catcacacac	acacacacac	acacacacac	acaaaattaa	ttaaaaggta	88680
ataaaagaag	cacgatcatg	cagagcggtt	tgctggctac	ttttacattt	ttacttttac	88740
attactcatc	tatgtttcat	tatttttatgt	gagttttttg	atctacttag	gcattttag	88800
gtcagggttt	tttaaatagtt	aaattcatct	acaatcaatc	ttactttatg	catatatgta	88860
tttatagata	cttagaatga	cattaaagat	tttggtagct	aaatctggcc	agtacaaaaa	88920
atatatatata	tctctagcct	aaaatggaat	gttgtcattt	atgagttcat	tgttacctga	88980
ctgtgggtatc	tgactaaaaa	aaattatcat	agtatagatt	gctatttttt	agaatgaatt	89040
atttttaagg	agcacttcca	tggaaccata	tctcagagtt	ggacaagagt	tctaaaaacac	89100
aatgtagaaa	aattgtcaat	attgtcagtt	atataatgatt	tccagtata	taacacaact	89160
aaataaagta	aaataaaaaa	atgatcccat	ttactggta	cttctaggaa	attcaggtag	89220
gtgtctgagg	tactgaaga	aactgggtcg	gttaccat	ggtgtaatct	aatctctttc	89280
aattctgtaa	gtctgaaata	tatgggtatt	tcaatgcttt	ttttttttga	gtattacatt	89340
aagtcaggaa	acagtctctt	ggcttgacgt	ttgagaggca	gtgcaaaaag	ccttctactt	89400
tcagtgtcat	ggcattgtcg	atcacaatat	ttctgatata	ttctgatata	agatctgaga	89460
aataaacaga	aaaagggtca	gtcatttcag	agaggtgatt	ccattataaa	gggtgggtatg	89520
tgtagcattt	tcttgagaaa	cttttaagtt	tttctttgat	ttttctctaa	gttttaagag	89580
ccctaagctg	ataaataactt	aatgaatttt	cagtaatagt	agacaacttc	catttagtatt	89640
aaataacctt	tgtgaatcaa	cttgtataaa	tcacaacaca	aatagtga	acaaaacaaa	89700

gcaacaaaaa	acgctaaact	aaactacaaa	gaagtcttga	gtgtttggat	ctgacccaaa	89760
gtgaccatta	aagttgtaaa	ttttctcatt	aaattgataa	caatgtaagc	cattctctac	89820
atcaattgaa	atgccaacgt	acttctgagt	aataacatcc	cttccatgtg	gtctctgata	89880
tttttgattct	agctttttac	gagcatgagt	atattcttta	ttttgaaggc	tgtgttgata	89940
gatggctctg	aattggaggg	ctaccaaact	ctaaagggat	tctttgctgc	ttgtcatggt	90000
tacagtcctt	tttctcttca	gtagcttttg	cttatctaaa	atgatgaact	ttctcagtag	90060
atgtgcaatt	caatcgagtc	atcttttccc	agatacagga	atgggttatta	atattctgcc	90120
ctctcagttc	tctatgatac	cctttatgac	cattcagaca	caggcattca	tacctcactg	90180
aggacagcag	ggacaggata	tgctgctgtg	ataacaactg	aaccaatgac	ttggcttcat	90240
gctctctaag	agctgccaac	tccagtgcgc	ccaataactt	ccaccccttc	actacttggt	90300
ccagtggtca	gatggtgggt	ggatggggca	ctgcaaggca	ctgctttctg	ctgttctgct	90360
ctaactcctg	accagatttc	acaaacagcg	tgtgctagct	ttaccatcca	caaagtgcga	90420
gaacttttgg	gggaataact	tgcacagccc	tcaaccagtt	ctttcctaaa	atgtatgtga	90480
agtcacccaa	cacattgaag	gttctcagta	gattttcttc	cttctgtctt	tcagtttggt	90540
tacctogcct	ctttgttttc	attgctttca	gattctttat	ttatagaagg	gaaaggggaa	90600
aagctgtata	ccacagccaa	aatttcaatt	ttgaaaaaat	tccaatctac	agaaaagttt	90660
caaatttata	gaaacaaaag	tataataaat	actagtaaat	acctttcaat	taaattcacc	90720
agttacaaat	tttgccacat	ttgctttttt	ctccctcatt	atctctatat	ctatccatct	90780
atctacctgg	ctacatctct	gtgtctgtct	ttgtgtctgt	atattacaca	tacacacaca	90840
gatgtgttat	ctttctatc	ctttctctag	tttactctag	aaatcatgac	aatatagatc	90900
atgtataaga	aaatcagtaa	taagttaata	ttatccaata	catagattta	cttatactgt	90960
aatttgtggt	ttcacogtag	tttgctttgc	tttgttttgt	ttctgatctc	tatattctct	91020
ttcatggaaa	gagaatgttt	atctgtgtgg	tgacttggtt	tatttttggt	tcogtaactg	91080
aatttctgaa	gaatccagag	gagttatcat	tagaatgttc	tgcatctctg	atttgtctca	91140
ttgtttcctt	ctgttttaga	gcaagtaaaa	catttctctt	gttaagcagc	aagtaaaaac	91200
tagcaagaat	actatgtagg	tgatattgtg	cacttaaagc	atttgaacag	gaggcacata	91260
attttatgtt	gctccattac	tgatgatgct	ggtttaatga	cttggttatg	atgggtgactg	91320
ccagggtgact	ccattatgaa	ggtgggtatg	tgtgtgtata	tatatatgat	gtaagagtgt	91380
gtatacatat	atatatgtgt	gtgtgtatat	atgtgtgtat	atatatgtgt	gtgacacacg	91440
tatatttata	cagtaaatat	atacatatgt	atatgtatca	cacaacatta	tatttatacg	91500
gtaaatatat	atactatat	gatacatgat	aagtgaagaa	ttgtggcctt	gtggcaatgt	91560
gagcatcctt	cttctctaaa	attttcatcc	agtagttttt	atcatctaata	aattggctcct	91620
tggtaaaaac	agttattatt	atgtattgaa	aaaaagatga	tatttctaata	tttgtcatat	91680
tcttcttatt	ttttaactgg	cattctttga	aaattacctt	tctctttata	tatatatata	91740
tttttttttt	catgattcta	agttatgtat	ttaaagtatt	ttaaatattt	taagcatctt	91800
gaaaaattta	aagtaggaca	tgttttataa	ttactgtgat	ttottatata	aacatcacat	91860
tttgataaaa	ataaaagagta	tattacataa	actctaaaaa	gtcttaaat	tatacacatt	91920
tctgtttcct	ttcaaagcaa	tttttgttaa	aggtaggaag	gagtaaaatc	cactgtgcct	91980
tgggttttca	ttttctattc	caactaagaa	gtgctgtcca	cagaaatgta	tcattcagcc	92040
actcagtcac	actttcctga	gaacctgcct	ccatcaaggg	acttcgttag	ctgtaaagtg	92100
ctgactaaac	ttgtgtttat	acctccactc	atgaaatagg	aaatatacac	acacatacac	92160
acatatatat	tcgtgtttat	acctccactc	atgaaatagg	aaatatacac	acacatacac	92220
acataggcat	tcgtgtttat	acctccactc	atgaaatagg	aaatatacac	acacatacac	92280
acacacatac	gcattcgtat	ttatacctcc	actcatgaaa	taggaaatat	ccacacacat	92340
acacacatat	gcattcgtgt	ttatacctcc	actcatgaaa	taggaaatat	acacacacac	92400
acacacatac	acacatatgc	atttgtgttt	atacctccac	tgatgaaata	ggaaatatag	92460
acacacacac	acacacatac	gcattcgtgt	ttatacatcc	acttatgaag	taggaaatac	92520
atacacacac	acacatacac	acatacgcac	tcgtgtttat	atctccactc	atgaaatagg	92580
aaatatacac	acacacatac	acacatacat	ttgtgttttt	acctccactc	atgaaatagg	92640
aaatatacac	acacacccat	ttgtgtttat	acctccactc	atgaaatagg	aaatatacac	92700
acacacatac	atatgcaacc	atgtttatac	ctccactcat	gaaataggaa	atacacacac	92760
acacacatac	atttgtgttt	ttacctccac	tcatgaaata	gaatatgtgc	atataatata	92820
tattatgtaa	ataaaattat	atatatttat	ataatatata	attttctaat	atgggctgca	92880
ggccaatata	catagatttc	tcctggctgc	aggtagccat	gtcacttgca	ttgattataa	92940
tatattttta	ttcatggaa	ttaaataaag	ttttaacatc	agagactttc	acagagattc	93000
agataacttg	agatagtaca	taatatgttt	aaaaagctaa	ttttttctg	tgaatattgc	93060
ttgtttccat	agacgtttca	ttctacaaa	gttcaaaaata	taattttaac	taacaaaacta	93120
tgaaggtaat	tgtaccatt	ctgcagctg	ttagcagctt	ttgattagat	tctcatagca	93180
gaattttccc	atatacagtt	gctaataatag	aagaaaaata	tcatctcctt	attggattat	93240
attgttaact	tgcttttgct	taatattttg	gcactctttg	gtaaatttct	ggaatgaata	93300
acaataaatt	acgcagaata	atctaataaa	ggatatttat	agcctatgtg	aaaataaaga	93360

tacgcatttt	agggactttt	gcagagatac	aaacacttgt	tctgtaaaat	atttgttaaa	93420
cgtagcacat	agggaaagca	cagatcctgg	agtgaanaac	ctacagcttg	aattccactc	93480
cagcatttat	tagaccagt	acttggtcaa	gttgcttagt	ttctctgaac	atcaattttac	93540
ttgtccataa	agagagagt	gtggtacata	atagatattt	tgacaggctg	ttaaaagcta	93600
gagatgaagc	gtgtcgggag	tctgatgcat	agtacttcta	ttaggcaagg	tttttcagag	93660
gaaaaagaac	caataggatg	tgtatgtata	catatgtaca	cacacacaca	cacacacaca	93720
cacacatatg	aatgagagac	atttaattta	aggaactggg	tcatgtgatc	atggggctgg	93780
caaactctgaa	atattcagga	caggcaggca	ggtctgaaga	caaggagggt	tttaacactg	93840
cgttccagt	ccaaaggcag	tctggaggca	gaattccttc	ttcctcaggg	aacctcagtc	93900
tttttctctt	aaggctttca	actgattaga	taagtacgc	acattataaa	ggacaacctg	93960
ctttaatgag	tctactgatt	tcgatgtgga	tctcatctga	aaaatacctt	cgcagaaaca	94020
cttagtctga	ggtttaacca	catttctggg	tgctacagat	aaaattagct	atcacagaac	94080
gtctacacag	aaaggtagct	acatttggtt	aaaccataga	attctattta	catgaccaac	94140
cagaatgtgt	ctagagggt	ccagtgcctc	tgcagatttt	ctatttggtc	tggcagtttc	94200
ttggatagcg	tctcctattg	gctgctctta	ctttctgtct	ttccattggt	tctttgaggt	94260
ttcataggaa	cagtctgtgt	ttagttaaaa	gtgcgctgg	ctcccagggt	cccaaagcca	94320
atgagtagat	aatgaatggc	aagcaatacc	tgatcccata	cagaagagg	aaaaccttc	94380
tctcccacac	ctgtgcctta	aggtagcact	gtttatttta	cagacttgca	aaaatctacc	94440
atgaaaacag	tgacttcaac	ttctaattta	gctttttctc	caattcatgc	tcaaaccaat	94500
ttctcaaatt	gagagcaag	gacgagaaat	gcttcatcag	ttaccttcac	tacagtctga	94560
tctctcgagg	tgcaactaaat	tcctttgtct	tcaactgatt	ctctcctctc	ccaacaggag	94620
atcaatgtac	agaggaatag	ttccttttaa	tccaggagt	catgcaacta	agcaacttta	94680
gagatgaatg	gttgcatatt	ccccggagcc	aattctagtt	gagctcaaat	gatataattct	94740
taaaaatagt	aattctattt	attttacaag	tgctatgtaa	acagagagg	aagtaccagc	94800
attataattc	actaaaaaga	cctgattttg	ctgcacaaac	ataactaaat	ggctacaccc	94860
tttaattgca	gaatccagat	ttcctccaaa	ctttatctca	ttccagtcag	gaaaacagct	94920
atcgtctttt	tttttttcat	caaagtcttt	tattatcatg	tttgaagctt	ttgattaagt	94980
tggcatcgta	agaaaccttt	tcctattttc	tttcatgttt	ttcttgacaa	gcctgagtoc	95040
gtcattcttg	ctggtattac	aggttcatgt	ttttgggtac	attcatttat	catggagtga	95100
gtcactttgc	tgttttttgt	tccactgtat	tctgaaaagc	acccacagaa	acatgacaag	95160
aatgcaggct	tgcaatggaag	aaattctcat	gccgggtctg	ctctgtctcc	ccagtcatgc	95220
tatcctgcat	aacgtcgtct	ttaaacaata	gaaagcaatg	gctaaggata	aaagaatggg	95280
ttcagataaa	actgaatttt	ctaccacctt	gaatattcac	tccatattta	atctgaagcc	95340
aaatatttta	acataaatgt	atctttattg	tctttgtcct	ggtcttccaa	tagggragtt	95400
tttgtagctg	gtacctcttt	gtttttttgt	ttgttggtt	ctactatttt	cttaatttta	95460
aattctaagc	ctccagggtg	cttgcttata	aaaacgattc	tggattttct	acttttaaaa	95520
tatggtttct	taaccatttg	cagttttgt	aagaggaaag	gagaatctat	aaagtatggt	95580
tacatgaaag	gaaaatctct	gtataacatt	ctatcaaaac	tgtaataaag	tgaaaattaa	95640
ttattaatat	tgatatgatt	tggctgttcc	ctccaaatct	catgttgaga	tgtgatcccc	95700
agtgttgagg	tgggagctct	ctgagagggt	ttcgggtcat	tgggatggat	ccctcatggc	95760
ttggtgctgt	ccttccccac	tctctctctg	tctcgtctct	gctgtatgac	gtgcctctct	95820
ccctttcgcc	ttctgccatg	agcaaaagct	ccctgaagct	tcccagagaag	ctgagcagat	95880
gctggtgcca	tgcttgtaaa	gtctgcagaa	ctgtgagcca	attaaacctc	tttcttttat	95940
aaactactca	gtcttaggta	ctgcttggtg	gcaatgcaac	aacagcctaa	tacaagtata	96000
cataatggaa	agtgaagcat	gccattttgt	cctttgaata	tatttggaca	attaaaccaa	96060
caataaaaaa	aagaaattat	atgcagtttc	tgactcgggt	tccttgaact	taagtggtag	96120
cagaagattc	ccctatcttg	aagagagaaa	gtcttcctag	gaaaaatgcag	gagctgtttc	96180
aagtgtgtaa	tggagcagca	ttcttctgaa	acacattttt	gcttgggaagc	ctaacaatata	96240
aataaaaaata	tgaagaggga	gcttttctat	tttaaggtgtg	agaagagggt	gtcgggtgcc	96300
cggttctacc	tcgttaagct	cctttcaaat	gctgagatgc	catcgacccc	tgattctgtc	96360
tcccaggaac	tcagactcaa	aactcttgca	ccagaaaaca	ctagatggag	aaaactacct	96420
gacaaaagtc	ctggacgatg	gccttgaat	caattccttc	atgtctttag	agggtctctg	96480
tagtttttcc	ataggtgctt	accagccaca	ttcttggagg	cgagctggct	aaagtaaaat	96540
tctgaactga	actgagcttg	atagagttct	gtgctcatat	tcccagcaca	gtctccatgt	96600
tgggagggaa	aattgctttt	tctcccattg	cagtaaaaaa	aaaaaaaaaa	agtcaatgat	96660
caagtctaga	ttccttacaa	cttctgttag	tttctctcac	tccatggata	atcaaatccc	96720
tcgcaacaaa	acagatgagc	tgtccctcaa	aattgttgac	ttgtgcctta	ttaggttggt	96780
agaagtgaga	atgtggtatt	ttaaattctc	aagactgact	tcgtgtatcc	atagagaata	96840
atgtatttta	tcctattttc	ttaaactctg	atgtatttga	ataagggcga	tttaaaagtt	96900
ctctcacatc	agtagaaaata	actgagctag	agtttgcttg	ttgacaatgt	ataggcacia	96960
tatatcatta	gaagagtgat	aatgactttt	attttggtga	aataataaaa	aaggaatcat	97020

aaaaatagag	ctgtctaata	atgaacccatg	tgttttaaaaa	agaaaagtta	actttctcat	97080
cactggaaat	aaggaatgag	ccagtggcca	tggatcacag	atataattgt	aaagattcct	97140
atattgagtg	ggaacacaga	caaaaataacc	cccaaagtct	ttgctaaatc	tatggtcata	97200
taaaagaaat	acagctggat	tgcctatcaa	aaagatccag	ggcagtgga	actaactttc	97260
tttacttgca	gccaaagtgt	tgaatcttta	gtagagcggt	ttgctttaaa	tttaatttat	97320
taacctgat	tactgcagtc	agcaaattaa	actcaagcct	gtattatgaa	atattatgaa	97380
ataaacactt	tagaactgtc	atctgtatta	agcagctgta	aaaaagcagt	gtcaagcctt	97440
gatgatttca	tataaaattg	aaataatttt	gaatattcct	catggtgatt	tgtatgtttt	97500
gttatgcaga	cacagaataa	aaacaagtga	acaaaaaaa	aatgttattt	aaaaataact	97560
aagaaaaaaa	acaatttggg	aaaaaatctc	gatatcggca	ctaatttaac	tttgtaaatt	97620
ttccaatgta	gttcatgttc	agtatatagg	cataatattg	attgtattag	aggggctcag	97680
tatttctttac	ggacttaagt	atgtttaagt	aagagcatat	ccattttgat	taatgagggtc	97740
tatatacctt	ctctcttgat	tactctccag	ttttggggaa	atcactttgg	tagctgcatt	97800
agctctttta	ggtcacctga	aaactcttta	ataaactaag	tggtaacta	gtaaacaaaa	97860
agaactcacg	cggcttccct	tgaacgaat	acatcactgg	gaattctgag	aaaccaatta	97920
aaacagaatt	attgttttga	atgaatcatc	agagaattct	agtctaaact	cttctaaaaa	97980
cctacatttc	tttctaattt	tattacatta	aattttaaga	tttttttcaa	actttttgaa	98040
tgagaatagt	catcagatta	ttttaaaaga	ttaggctaaa	atatgcaaaa	ttcccttcag	98100
taatacattt	ttttaactgg	aattataatt	agtcataagg	ggaaagtcat	tttagtgagg	98160
ataaaaattc	aaaaaagcta	aacattaaat	gctgaagttg	ctttttgccc	tttgcaaat	98220
ctttcttcat	tatcatcatt	tctctatatt	tttgaatgca	tatatctcca	aagaaattac	98280
tttttgctta	atataaaatt	ttaggcaatt	agatgcaaaa	gtgtagtaaa	gcaacactaa	98340
agttgccaat	ttaaacatag	ggaagcaaaa	cagcccagga	gttaaagttc	tgacagctat	98400
cttaactagc	acacatggcc	agtgccactc	agtaacacaga	cacacaccca	ttttcacagt	98460
cactgaggag	ctctaagagt	tgattttgaa	acgtagtggg	aattaccttt	tttgggaaaa	98520
taacataaat	ctgagataaa	ttcttgactg	tccgtactca	aaatattttt	ttccaaaatg	98580
aaaaaaatgc	aaatgtcctc	agcataataa	tctaggataa	gtagttaatg	atgctacatt	98640
aaggcattac	ttggggcttg	caaattcatg	actgacattt	gatgatattc	cagggttagaa	98700
gagttatttc	ctttctttct	ttcctttctt	ccctccttcc	ttccttccct	tctttctctt	98760
tctttctttt	tctttctttc	tttcttctt	tttttatttc	ttaaaggaatg	tgtaaattaa	98820
taacaaaaac	cttttaacata	taaacagatt	gataaattac	tgagcaaacc	aaactgtcca	98880
cccatgtttt	tccttggtgt	agtcgtctcc	aaccatacca	gcaaggccta	gaaagagcaa	98940
aaaaaaaaaa	aaatgtatat	atgtggcggt	gtgatatttt	cttaaacaaa	gattttttat	99000
gtaatgccta	actatttctc	tcatttgaat	ctgagttgca	atttagacac	aaatcctgag	99060
gtcagatcct	ggcttatagc	tggatactat	agagcatgtg	ctgtgttcc	gactcctcat	99120
ctatgatgta	gaaatgataa	atcccacagc	agaatgttca	agtgaaaaag	aaataggaaa	99180
atataacgta	tcatgcataa	ggcctcttat	aaaaaacaca	gattccaaaa	tgacatggtg	99240
attatcatca	ttttattata	gccaataaat	tgccagggat	tttaaaaact	atcattctct	99300
aagaattcta	tttaagaaca	tacatataga	taaaatatgc	attcaagatt	ttgtaacagg	99360
atctggaaat	aaagtcaaca	caaaaacttt	ctatcctgct	gctgggtatt	aattttctat	99420
tcttgaacag	atattaattg	tccacattcc	ctactttctc	tctcaagttt	acttaaaaaa	99480
atatctactg	ggattatcta	catgggtgca	gacacccgtg	aaggataggg	tagtggagaa	99540
aataaaacatg	ctcccgtttc	tatggtgttt	atcatataat	ggaggacatt	cttgggttgt	99600
gcattttatt	ggttcagcta	tgtaagggat	ggtaaaggga	atatgttcat	tacagaaaca	99660
tattcttact	ttctctccta	attacagtaa	atgttcagag	gtatgggaat	ggtcaaacag	99720
aaaatggaaa	attataataa	aaaataagaa	agttatccaa	gtccacagta	aaactataca	99780
atlttataac	acagaccaag	aaaatatctg	catagaattt	ggtatgcttt	ccttgcccca	99840
tcaaatccac	catttttagtg	ctctatattc	tctttctgtt	tttaacatat	gtaacggcgg	99900
ggctttctga	gctggcttcc	ttggagtgca	gacattagta	atacagacaa	cttctattcc	99960
tcttccccct	acacattgct	tgtaaatcag	ttattattta	tttatatgtt	tttaattttt	100020
tgaattatga	cagagaagga	gtagatgaat	ttaggtagaa	agatagggaag	gaattgctga	100080
gccaaaaact	aacaaagaaa	ctgggtcttg	catgggggtta	gtctccaact	cagaataaaa	100140
tgaattttta	tataagatga	ctgttcagat	caggaggata	gaaattaaaag	ctagaagccc	100200
tcaaaagttg	gagaatataa	tagaagatta	taccaaaatt	aagcagaaat	tcaggatctc	100260
ctttcaggta	tgtttgaatc	aaaagcaaac	cagggtattg	caacacaaga	tacacattgc	100320
cttgggacgt	tcagctaccc	tcaaaccatg	aattttgctt	aagggtctca	ggttggtaat	100380
atcactaacc	tgagtgtgac	agagacaagt	caatcaatct	ctggaataag	gcaatttcat	100440
cataggcttc	agcttattca	aatcagtttt	caaacatgat	gggcagcata	tactacaggt	100500
tcaccaagca	taggcagtga	tttatctgtg	aatgggaata	aacagaaata	acccattaag	100560
tttacagata	ttgaaatttg	tagacacagt	ctacaaaaat	actatgtttt	tagggaaaaa	100620
ggtaagtata	taaactcaaa	caagggaacta	tgtattaggt	tatctaaaaa	atttaaaaaa	100680

gatcaaagga	aacttctata	aataaaatct	aaaaacctga	aaccataaag	cctatgggtg	100740
aaattaacag	tacattagaa	ggagatgaag	atagaattgag	cgagttagaa	gaaactcaaa	100800
gaaattatatac	agaataaaac	agaggaatta	agagacaagg	aggacagaat	gaatggttag	100860
gctttatatt	gaatcagagt	ccttgaaaaa	aaagaaaaaa	ggatgggttg	aaataatgtt	100920
tgaagaaata	gtgcctgaga	atattccaga	gctgaggaaa	gacacctgca	ctcagattca	100980
atgattggaa	caatttttca	gtagtataaa	tacaaagaaa	tggtgggtgat	tgtaaaactg	101040
caaaacacca	aaatcaactg	aaaattttta	atgtaacctt	gaaaagaaaag	agaacataag	101100
aaattagcat	gataggagaa	atatttaaat	cagtaattga	tttctcatca	acagcaacgt	101160
aagatgaaag	agaatgaaat	aacttcacaa	gctgaaacat	aataactgaa	tctaattgctt	101220
gaatttcttt	cacaattata	ttgtctaata	atccaacaat	cacattagaa	attaaatagg	101280
agcaagaata	cttactgcag	ccttatttgt	attgctaata	gcatagaaat	agcatccagg	101340
tcaataagta	ggcatttcga	taataattat	gagatagcag	ttcattgaaa	tggtgggcat	101400
ccattagaaa	ggaaaaggca	attcttaaaa	atccacgtgg	agttgtattc	cagatgtaat	101460
ttttgcaagt	gcatagccat	ttgcatacag	aaaatgtatg	tgatgtgccc	gacacatgta	101520
tagattttgt	ggtaattctt	ggagagatac	tcaagaaaat	aattgtcgtt	gtggagagaa	101580
atgaggtgct	gaagaaaagc	agccaagggt	ggggggattc	atactttttc	tacccttgt	101640
accttttaaa	tctgttttgc	catgtactat	tattcatgaa	ttacatatat	tttttaaggt	101700
tggtttgatt	ttagttttta	attaaaaaca	actaaaatac	aaatatctcc	acagttatca	101760
acagcacaga	ttaaggaata	gtgtgttgaa	gtgtcttcat	tttaggggtg	attgtgttaa	101820
agttctgggg	acttaaatga	cctttttcat	accatccaaa	taccgattga	gaatgtaaat	101880
ccaaatctct	ctacaaaatc	catcctttta	tttttactgt	ttaatacctc	cacatatttt	101940
ttaaacaggt	ctatacattt	ggtactatat	aagttgtatt	gccatcaaat	gaatattcat	102000
gttcaactcca	acattcatac	gtgtttttca	acctatgggg	tcctgggtag	atgattaggt	102060
catgagggca	gagtcctcat	gaatgaggtg	agtgcctca	tataggagac	cccagaacac	102120
tcccttgctg	cgctcctcca	ccatgtgaag	gccagcgaag	aggaaggcca	tctatgaacc	102180
aggaagtgat	ccctctcctc	tccagacacc	aagtgaccct	gcaccttgat	cttgggcctc	102240
gtaacttcca	gaatcttgaa	tgtatatctg	tttataagcc	aaccagtctc	tggtattcca	102300
ttatagcggc	ccatacatac	caaaagctgt	ggcactaaca	aaatccagca	aagaagagtt	102360
ggatctaatt	gtttcattaa	aatttgatta	tgaatctttg	ctacggccta	aacagcagag	102420
accagagttc	tctccttcca	ctgtgcgagc	ttacaaattg	gccgctaaga	gacaacaaat	102480
ctaccagcac	cttgatcttg	gaattcccag	cctccagagc	tgtaagcaat	aaatttctgt	102540
tgtttatcag	ccaccaatt	tatgcttttt	cattagaatg	gccccaaaaca	agtaagaaaa	102600
tccccaaaaa	caaatacaag	gactggtaca	tgccatatag	ctcagtaata	tttttttgaa	102660
tgaataagtg	aataaaggta	gagtgaatgt	gtattgcctt	tgaatgtaaa	cctttaatac	102720
caattgcagt	gtactactga	attcaactaa	gaattatttt	tacatagttg	tgtagaagca	102780
atttataata	tggaaatatt	tctcaaattt	acagaaattt	gataaatatt	atattttatt	102840
tatgtaaatt	cccaaactaa	aaagacatgt	cctttctctc	actgtccatg	taagttagtg	102900
gaagacatcg	atctcaaat	acaatattta	actgtattaa	attgtatata	tagatgcaga	102960
ctctaggata	aaagattaaa	tgtttgatat	ttttaagcca	gatccotatta	ttaaaatcat	103020
caatgatttt	tactgtgaag	gtgtagccac	atgcagttac	cttcagaaaa	cttcaaagca	103080
gcaaaaagcat	gtcgcagagc	caaatgcagt	tccagtgccc	aagaaaatag	ctaataatac	103140
tgggcatggg	tccaaaaaaa	cccatatggc	aattgtagtt	gaacatcttt	tacagcagca	103200
attccattct	aacccaagtt	ccttccctta	aatagcacct	cttttatttc	ctgtagtcaa	103260
actggctgta	cagtcctcac	tcttgccaaa	gtccttgtct	gggatgtgca	tctccaggaa	103320
ccgaagtgt	aaacettctt	ttttactagc	ctgctctgcc	aaaatccatc	ttgacacacc	103380
ttaaattata	ttagatcaaa	ggctgaaaat	tatatgccat	caaaagcaga	aagataaaat	103440
gtttatctac	cgcattaggg	atgtgcaatg	tttttaagac	acttgatggt	tgagtggagc	103500
attattcgac	gtctctgcag	tgacgcatgg	cttacttctt	tgagtaggtt	gatagatata	103560
ttgtaacagg	ctggtggcaa	atatgaagtg	gcatattcag	tatgtgttta	ggatgtttat	103620
gtgttctctg	attcttgttg	tgtctgtctt	aattggaaat	aaggtttacg	tggtgagagc	103680
agtgaagcaa	cgacagggca	atactataaa	agcgaactct	gccaccagca	ctccatttct	103740
tttaattgat	ttctacatat	ttggaaaact	ttctaatttc	tccttaaaaa	ataaaaccat	103800
aaatgtcaag	aagaatgttt	ttcttcttaa	tgaggtttgc	ttgccaaagc	ggctttgttg	103860
caatgtctgt	aggtgggcta	cggtaaagac	cagtgggaatt	gaagtaattt	ttaaaggaat	103920
attgtctcaa	atgaaatata	aacaagggaac	tgtgtaccca	ggcttcaaag	tgataaacag	103980
atgatgctat	aatcaagtaa	tttttaataa	atatattgat	gaatttactg	aagtccatat	104040
ttcagtaatc	catcatccac	tccagatat	acattgagat	tacaaaatta	acaatttgtg	104100
ctgttgccag	aattaaaatt	tcctagtaaa	gttgaactat	tatataacct	tagtattttt	104160
gcacccctaa	ttttttccag	attattttgc	aggctataaa	catgttaacc	tttaggacta	104220
ttttatgtcc	aggtactact	taaaaatgta	tttttagatac	aacaatcata	caatattttc	104280
taccgttatg	gtttcacagt	agataactaa	aattatatat	atatatatat	tttttgagac	104340

aaattctcat	gctgtcacc	aggctggagt	gcagtgggtat	gatctcggct	caactgcaacc	104400
tccacctccc	tggttcaagc	gatttctcctt	gatttctcctg	cctcagcctc	cagactagtt	104460
aggattacag	gcacctgcca	ccatgcctgg	ctaactttog	tattttta	aaagacaggg	104520
tttcgccatg	ttggccaggc	tttaacataa	ttgattatgt	ttcaaaaaat	aagaagcaat	104580
ttctgcatac	aaatttcaga	aaaattcact	tggaataag	ttttattagc	ttacactgtg	104640
ggtagagaaa	agtaaaacat	tgaaaaggta	actgtttgct	ttaacattta	cctgtaaaaa	104700
cattcatgaa	tttaacctaca	atttcttggg	agtagaaata	tattctcaat	gttgttttag	104760
aagcacatat	ttaatggaga	aatttgtttt	ttatagtgtt	atttttaaac	ataattttta	104820
atttagaaaa	ttatgacaaa	aactcgttca	ttcccttttt	tggaacaaac	atacttaacc	104880
actagtaaat	gttaagtcct	gcattatgca	caggggttaa	taaaatttag	gatgacatgg	104940
ctcaggcttg	tttaagacaa	atagcactga	taggctgttc	cataaaactt	gttatattgc	105000
ttggaagaaa	tgaggcccat	atataattga	atatatcaaa	gttagtaaaa	agacaaccac	105060
cacgtgttta	tccagccctt	gcaatatagt	gcattctgtg	ctacactgtg	aagatcgcta	105120
accattatga	agattctttac	agtcattggga	tgacaagact	tctgtaatgc	aatgattaca	105180
gcactgtgtg	atgaatgctc	tgccctgtgt	ggataaaaatg	ctttagaaac	atgagagata	105240
cctgcaaaaag	gtagctagta	aaagcttcag	agaagatagg	acttctgaat	ttgggtctta	105300
acattctaaa	aaaagaaata	gtttgaccaa	agacatggaa	atttatttga	atttgacata	105360
ttttgtttgc	ttgtttgttt	gtttgacaga	gatagggttt	ctccatgttg	ctcaggttag	105420
tcttgaactc	ctgagcaatc	tgccctgtctt	ggccctccaa	agtgtctggga	ttataggtgt	105480
gagccaccat	gcattgtctt	gaatttgaca	tattttgata	tttgggtaag	gggagaagat	105540
tagtccaatg	gtttgcacat	tggaacgctc	caaacctcat	gctaaaaattt	gatccccagt	105600
gttggaggtg	gggcctagtg	ggaggtgttt	gggtcatgag	gggtgattct	tcatgaatga	105660
cttgggtccca	ccctggcctt	aatgagtga	ttctcacaat	gtcagttctt	ggaagtgtg	105720
gttgttaaaa	agagcctggc	acctcacccc	tctctcttgc	attctctctg	cacacatcag	105780
cttcccttgc	ccttccacca	tcattgtagg	ttctctgaga	tcctcaccag	aggcagatgc	105840
tggtatgctt	cttgtacagc	ctgcagaact	gtgagccaaa	taaacctctt	tcctttgtta	105900
attaccacag	ctcatgtatt	tctttattga	aacataaatg	gactaaaaca	tttgatgtgg	105960
ttgaagtgtg	agttttatgg	aagagattgc	caaggatgtg	tccaaatggg	gtgggtgggtg	106020
aaagactttg	aacgttttcc	aaatgtatgc	caatacatct	agactttact	tttgagctag	106080
aggaagaaca	ggtgcagcca	aaattaattg	gaaaagtgtg	agattgcatt	ttgtttgaga	106140
aggataattc	tgctattaaa	agagttagat	gagtcaagaa	gacaccagt	gacaccagt	106200
tcagtgtgtc	ttgtgaaaga	ggactttggt	tcattggagct	gggggaaaga	ggtcagagac	106260
catgggcagg	attttgtggt	caacttgaca	tcgggatgat	gtacaaagaa	aaggcaagaa	106320
ggaacaatct	gggcagtgat	gtcagtgcaa	atttgggcag	atcgttttga	tagaatggat	106380
caaattggaat	ccaaataaag	taaaattctc	aattatgaga	attcaattaa	ataattttgc	106440
tttaaaggaa	actaagattt	ttaaatgcca	aagataaaag	ctgaaagtca	ttataacttt	106500
agtcttccct	aaaagtgtact	taagcaagaa	aaaaaaacag	tataattttg	aaactagtta	106560
acagggtatt	tggtttaatt	taaatgatct	ctccagcaga	tattcctttt	gtgggtccata	106620
aattttcagt	tggtttattta	tttttgacc	atacagtaaa	gggttgga	gagataaaaa	106680
tcagggaaaa	ggaaagtcca	gacacagaaa	ccgctgagtg	ggcagaagca	ctgtccgaga	106740
tcaggggggc	acaatgatgg	aagtgcagtt	aaggttggtt	ccacgctgtc	ctagatgtac	106800
agagcaaaag	gggtgagct	ctcaaagcca	cagttagaaa	ggaggtcact	tgagtgcag	106860
gagagaaatg	ctggcaagag	aactgggcca	caaacacagt	ccaaaggag	atcacggaag	106920
cccatgagga	aagtcaggaa	ggagtactg	caactcaata	gctgctatca	taaaaacaac	106980
cctgctgcta	cgtgggagct	ggtgtcttcc	tgtaaaatat	actccaccta	attatcctac	107040
agcaagatac	tgcatcagtg	cattcccatt	gccttgtaag	tccgccttgc	cttgtcttgc	107100
cttgccaagg	gtgtgatcca	aaactgagtg	ggctgatatg	aattgggtgg	tgctttcaat	107160
tcctacata	gtttgcctcc	tgaaaacctc	ctttggtatt	tgggccacgg	taatgcaaag	107220
gagccattca	tccaggtaaa	ttggagctgg	aggtaggggc	tgagcctctt	ctggggcctc	107280
rgggggcttc	atcatgggtc	ctggcctgag	tgcaaagata	gcatctagta	tcacagataa	107340
cttctttttg	atgcctagcc	atatttctga	ttttcttgtt	tattaagagt	gagaactaca	107400
tttctacta	ggctaggtct	tacttcgcag	ttcttaaaat	gtgttaaac	tatttgtggt	107460
tctcctttga	attcttccca	agttctctcc	acgcttgagg	agttgtggag	ctggctcctt	107520
agagatgcta	tgtctatgat	gccaaaccca	ttacaaaaca	caagcaaaaa	caaacagcta	107580
cgctagactt	tccaaactct	aaaacatcga	atctcctctt	tgaccctatc	atgaaagtga	107640
gttacttaaa	ttaaagagta	attgtatttg	tgtaaaatac	attattttgc	aaataatttg	107700
actgtataat	cccagtgtag	ttataattct	gtgtggctaa	ccctgtgttg	ctctggaatt	107760
ctgtcctctg	attttatgaa	ttctaagatg	ctgatitaa	ttaatttttt	aaaatactct	107820
ccagtttact	tgagtccaag	taggctgac	cacagatca	tattcaaaaga	attccaatga	107880
gctgttgagg	aggattttta	gcctccact	agcttctctg	catcagttgg	ctttgagaac	107940
caacctaaagt	atttcagagc	tgctagtga	atgaatcttt	cttacagata	ttacagaggg	108000



ttacagaaag	ttcagtttct	gactcaccct	gattttgaga	aactgtttct	tagggtcatt	108060
ccctaaggtt	ggaatcgctc	tctctctctc	tggttggttag	tgccctgtaa	acactgacaa	108120
ttttctttct	cttcagttct	tcattcttgg	tctagtgtct	cagtttccaa	gtctacatga	108180
ctatgctaca	gtgtaagggt	atttagacgt	acaatttcat	gatttctttc	ttctgggatt	108240
atttgcttga	cttgtgatta	gcttgccctga	tcacagacaa	tagtaacttt	ttaaccattc	108300
ttacatgatt	tcctttttcc	ttaaatttgt	tgtttctgta	ttagaagcag	gaaaaaagg	108360
ccaaagttta	gaatttgaaa	agcagtgttt	gaatgcactt	gttacctggt	aaaccagtca	108420
tttactgcta	tcttacaat	gaccacagca	catggtagac	ttaaagaaca	attatttgca	108480
cattttctca	ttttgtggat	cagaaactgg	cagcactcag	ctgagagact	cttctgtttt	108540
acggggattc	acagcgttca	ctcagagaca	gatagctggt	gggcaggccg	gtgtcaggga	108600
ccaaggatgg	catcaactcg	gtgtctggcc	cttgacagg	atgctggagg	tctgggctca	108660
cctgggaatg	gagatggcaa	caactacatg	tgacctttcc	agcatggcag	tctcagagt	108720
gatatggcaa	cagctgcaca	tgacctctcc	agcatggcag	tctcagagt	gatatggcaa	108780
cagctgcaca	tgacctctcc	ggcatggcag	tctcagantg	gagatggcaa	cacctacatg	108840
tgacctctcc	agcatggcag	tctcagagt	gatatggcaa	cacctacaca	cgacctctcc	108900
ggcatggcag	tctcagagt	gatatggcaa	cagctgcaca	tgacctctcc	ggcatggcag	108960
tctcagagt	gatatggcaa	cagctacaca	tgacctcacc	agcatggcag	tctcaggtta	109020
ctggtacttc	ttaggtctca	gctgaggact	ccagtgcagg	tggtgcaaga	cagcagacgt	109080
agcactgctt	atgtgtatcc	tgaagtccca	tggttatcct	gaagtccag	aaagctgctt	109140
caagtgcatt	ccagtggaga	agggagtcac	caaggctgga	cagtcattta	tccatggcag	109200
gactagcaaa	aaaaatctga	ccatctctca	tcgcacacac	gtgggtaagg	tttgactgg	109260
tatgatcttt	ttccataatc	tcattttcgt	gatttcttaa	ctgagaatat	aatagtaaat	109320
accaatctca	cagaattggt	atgaggaatg	cagaaactta	atcaaattac	ttaattgtta	109380
ttttcatttc	ttattaattt	gttcccattg	aaattaatgt	tatgtctggt	ctcattttaa	109440
attattttatc	tttctgaatt	tgatcccat	cctcaagttt	tcaatgcttg	actttttatg	109500
aattcttgta	ttatacctca	aatttaataa	aatcctttta	atcttgtctt	caccctagca	109560
ctgtctatac	tgtcccctgg	gtggaagttt	cccattcttc	taattacttt	tatacacaaa	109620
tttgacttct	ctctccagcc	cctgtgagaa	acatgtcctt	tctgttacac	aggagtagg	109680
tgaaggcctt	tggaacttct	atttttccta	tatgaaactc	tcttctttca	gatatacaga	109740
ttttctctct	tcttttcat	tcagaactct	gctagacaat	catcttacca	ggaagatttt	109800
tattcatcac	cctgtgctaa	ataacattgg	tactatgctg	ggctgagcta	ctatttttcc	109860
ttgaatcctg	ctgtaggctt	ttcttcagac	acttaccatc	accaggcgta	atgtgaatgt	109920
gtttgttcat	tgcatgtott	ctocacgtga	atggaactt	catgacaaga	atgtgttcca	109980
gggctggg	ccgcggctca	cgtctggaat	cccagcacat	tgaggaggccg	aggcggg	110040
atcacaaggt	caggagctca	agaccagcct	gaccaatgta	gtgaaacccc	gtctctacta	110100
aaaatacaaa	aattagcctg	gcttctgtaa	tcccagctac	tcggaggctg	aggcagtaga	110160
attgcttgaa	ccctagaggt	ggaggttgca	gtgagccgag	atcactcaag	aaaaaaaaaa	110220
agaaaaaat	gttttccttc	acctctgaat	ccctagtgtg	taggggagag	cctggaacag	110280
agttaggggt	ccataagcca	atcaataaaa	tgactgaaaa	agagtttaag	gacaaggtca	110340
aaacaaaaga	gctaatttaa	ctgctaaacg	attattctat	ttaggccaat	tggttaacct	110400
tgactgcttg	taaaattcaa	tttcattgaa	tgggacatat	ttcaaacata	aacataggtt	110460
tttttaatga	atataactga	actgcagatt	gcaatcaaga	tgttccctcc	ctttctcatt	110520
tagatttctc	atcttattag	ccatcactga	ccctctctca	gtgggttaaat	gcttagtttg	110580
ggacattatc	tcagccatac	cctcagaatg	gttgattcag	tggttcgttg	cccatgtgtg	110640
ctgtgatgtc	aacaggctgt	tgaatgctga	gggcaatctc	atatttggtt	aatgatggtt	110700
cttcagaaa	caacaccctt	cagattagtt	cagcagggtt	caactttgtg	gtgaccacag	110760
ttgggctgag	caggctgcat	gtcaggcaga	gctgtcagtc	tgctccagcc	aagagtggag	110820
tgggcacttc	ttgggatttg	gataaggaaa	catttgtctg	agcagtttta	ccaccttaat	110880
atgggttga	ttcttctatt	tgtatacaga	gtaagccaaa	ttaaaaaatg	aaaagaaaaa	110940
agaaccagtt	catttaataca	agacatgcta	gtgggttagtc	atcttttata	acaaccacat	111000
cgataattca	cctttttcat	tttcaaatga	gttttggtt	gttttgtaga	cagagagcct	111060
tagccataaa	ttaaaagtaa	tgataattag	cccatgttcc	ttcattatgt	tgagggtttt	111120
gaacatttct	tcacatagct	attagacatg	tcagccacat	gtaggaaaag	gaactataga	111180
atttcgtaat	tttttttttt	tttggaaaaa	tgagcaacag	aattttaagg	cagaaacgtg	111240
aaatgcgagt	gatggcaagt	cccagtgaag	tggagaatag	catgctttat	gtgaatgtct	111300
gatatttggt	ttctcagtgc	ttttgatctc	agaacaccta	acaggactgg	ggcagaagct	111360
ttttgtctac	ttgttgcatc	agtgctaaag	atgaaaggaa	acagcttcca	tattcgccgt	111420
ggtgttgatt	gccctataaa	taggaaacta	cccagttaac	cttgccctgtt	gtcatacggg	111480
gagttttatt	aaaaaataaa	cagtgcgtgg	cggttagcag	aatgtgggat	ttctgcttag	111540
atgcccga	ttctttcata	tacttcatta	aaatgtgcag	gagaacttca	gagccatgaa	111600
taaaaatatt	gttatggaaa	ctgaggcctg	ttatcatagc	ttcagtgtat	tgcaatgatt	111660



atttaatgaa tgagtttatg tttgaaaata tcccctgtga taatatgaat ttatcgctcc 111720  
 cccctacccc tccgatttct tttatttttga cacatatgga agatgctgtt ttgttttgac 111780  
 actcagagga gatgctcttt ttatcaaaaa tcaactgaaca gtaatgagat tttagtgaac 111840  
 aatctttacca gacttccoga acataaagaa atctgatact gaggagtcca cagtgttcta 111900  
 taataagtaa tcttaggtgt gtgcatgctt ttcagtggca gtattcattt ttcacgagtt 111960  
 ttgtatttga tcttaccagg atttgatgta agaagacttg ggtagctctt gtgtagacac 112020  
 aaagaacaga gaggacatcc ataaaagccat agagacagga tcccaaccac agtgtaggga 112080  
 gaaactgatg actatgccag gcatacagca ccatctcttc agcctaagag caggggggtt 112140  
 accaagctca ccaaaacaac ttaccattcc tctttccgcg tgattggaac tcatcttcc 112200  
 cctcagaggg tgcagagatg aatttagaag gaagtttgaa gagcccatgg tgtccagaac 112260  
 tcaggcctgc ttgcaaatga gtggctctgg gactgtcttt ccacatctcc aggagcctat 112320  
 tttccgggag gaaaatgaga gcaactgttca gaagtactgg aaggaacctg ccaacattga 112380  
 gtttcttgag agtctctgaa gtttagttctt ctaattttta gtaccctgac aatttcattg 112440  
 aagttagtaa cttttgcata cttcagtttt ctcactgtga aaattgggtt aatgataata 112500  
 tctaccaccc tcacagacac acccaggatt aatactttgc atccttcaat ccaatcaagt 112560  
 tgatactcag tattaactat cacatggtcc ctgaaaaattt tattttattc tgatttggga 112620  
 ataaaaaagc tttctctttg accccgaaag tttttctctt tttgttaaaa tatattaata 112680  
 tatttatatg gatttctaaa atattgtgag cccaggaac tctggttaat gtgcctgatg 112740  
 atctctgttg ccgtcagggt tggtttctca ttagggcaact tctgggcttg tagagggcca 112800  
 ctttctcacc ggttccaga atggaccttc caagatctct ggtctcctt cctcttaaaa 112860  
 ggacacaggt cctatcagat tagagtcctt cttctatgac ttcatttagc ctttaattatc 112920  
 gccttaaggg ccctatctcc aattgccttt ggggctagga aatcaacata tgaatcttag 112980  
 agggacacaa ttaactcaat aacagtgaaca ttttaaaac tacaaaataa gggaaatatc 113040  
 atagtttttt tctttacttt cacatagata ttacatcata ttaaaagata gatcatttct 113100  
 tcattttaat tgaaaccaga atttatattg aataagggtt gaccaatatg ggtttttcaa 113160  
 catttactct tatattaatg tattagcaga agacaagagt tcatatttc ttaataattt 113220  
 tcctagaaat tgactaatat ttttattttt actttctcag atatatctgt ttaataaac 113280  
 tatttttggt gttatcta atgtattcaat tttagacaatt tggagaacat ggatgcctat 113340  
 aaacatacat aagaatacaa atgtatcctg ctttcagcat cagagataat ttagggttat 113400  
 tgcaattggt tctgttttg caaattggaa ggattatggt tgatgaactg aggttaagtat 113460  
 ttaggaaga atatttgctg ttgggtcagg catgctttca gtttttatgt atgtagagt 113520  
 tgatagacca ctgagacatt taaaagatgt tcagaaggta ttgaactttt ggaaaggcag 113580  
 atcaaaatcc cacaaggga gctgacctg tgtacagaca tttgtgaatt gtctgcatat 113640  
 aggtggtatt gaagccatgg atgaaattgc taactaaagc atactcttgt gagaaaagaa 113700  
 gaggacaact taccatcgg gcacattaac cacagtctct ggggtctaca atattttaga 113760  
 aatccatata aatatttta tatatttta caaaaagaa aagaaaaact ttcagggtca 113820  
 aagaaaagc ttttttattc tcaaatcaaa ataaatttaa attttcagaa accatgtgat 113880  
 agttaatact gagtgtcaac ttgattggat tgaaggatgc aaagtattga tcctgactgt 113940  
 gtctgtgagg gtgtagcaa aggagattga cattgagtca gtggactcag aaaggcagac 114000  
 ccacccttaa tctggctggg caccatctaa tcagctgtca gcatggccag gatattaagc 114060  
 aggcagaaca ccgtgaaaag cctagactgg cctagcctcc cagcctacat ctctctcact 114120  
 ctgaaggctt cctgacctca aacatcggac tccaagtcca tcatcttttg gacttggagt 114180  
 ggctctcctt tctctcaac ttgcagatgc cctattgtgg gaccttggga tcgtgtgagt 114240  
 tcatacttaa atacttaata aactcctata tacatatata tatatatata tatatacata 114300  
 tatatatata tatatatcta ccatattcgt tctgtcactc tagagaactc taatacatat 114360  
 aataaaaaatc tattaatttt ttgtctgaaa cagaaaaaat aaaatattta acctttaaaa 114420  
 atgttcaatg ttgatatcaa tatgatgcca gagcccaatg aaaacaaaag tacttagttc 114480  
 ccacaaaagt tataatgaga cctgatttg acctagtgtt gaaatagcaa cctttaaaag 114540  
 gtttgagaaa aataaatacc tacacaagt tttgaaaata tcccaggtgt ctaacaactc 114600  
 actttctcta tgtctttccc atacccaaa tatcaaccat gacatttttc cgctgatctt 114660  
 tcagatctca gaaggcgaca tgaaggagat ccctctgctg tgagattgct attttcctat 114720  
 tgtacctcaa ttacattttat aaacttcttt ctttcoatgoc agcctaagtg gtttttgaat 114780  
 ttattaactt taatgtttta taaaccagtt tctatagcaa caatcctatg acagatgttt 114840  
 tgttaaagtc tctttaaagc acacacctct gtcaatatgt tattttgaac atttcaaagg 114900  
 ataaactctt gaaaaaggct gcacatatct caactttcaa gctcacaatt tgcattaata 114960  
 attataccct ctcttaataa aagaattatt taatagaagg ctgatgagct aattttaaga 115020  
 ataatacagt aaaaggtcaa aataaaactt accactctcg agaagaaagt gtaagggtcaa 115080  
 gaaaataaaa aaaattacaa catggactat actactttcc tcaatgttaa ttactttaat 115140  
 aaatgtgtga cgaatgatta ccaacatttt aaaatctatg tgcatttttc tccaatgagt 115200  
 ggcactgggt tctttgcac ttttagagt ctcagtcctg gcatatgtag gtcactgacc 115260  
 gggaaattac catctcaatt cactttcttt cttctttgct atttcttato tgatctcgaa 115320

gttaaggcat tgaaactatt gcaaaaccct cattatTTTT tcccaaatta aaatggtaat 115380  
 atttaatatgc tgcataTTTT gatgcattgt tttggctctg gtttccctttt tttatgtgtt 115440  
 tttttttatt ttgtgtttta tgaagggttt ttttttattg cttttttgtt aaagaacgta 115500  
 gcaaaaaaaa tgaagtatac taaaagctat ttcaccaga gaaacataat ttgttgggtt 115560  
 tttaaaaata tcattttcat ctgcttcttt gggaaacatc agaaataggc actgagggat 115620  
 ggcaagggtt atgtgaact ctaagtaatg catgttgtca cttgcagcca gtcactgaac 115680  
 tttctctgcc gttagcttccf catttttaat gtggactaag actccacttt tatatcacca 115740  
 cgtccaagtt tagacacgtg accctaagtt gctctatcag ggatctctag gccttcctta 115800  
 tgtgagacac aagtggtgta tggactgcag tcaaaaggcaa gtgaggtttt tttttgtttt 115860  
 tttgtttgtt ttttcaattt ttgagatgga gtctcgtctt ttcaccaagg ctggagtaca 115920  
 gtggcgtgat ctgagctcac tgcaaaactc gcctctgggt tcaagtgtt cccctgcctc 115980  
 ggcttcctga gttagctggga ttataggtgc atgacatcat gcaactgctaa tttttgtatt 116040  
 ttttagtagag acgggggttt accatgttgg tcaggctgggt cttgaactcc tgaccttgcg 116100  
 atccaccgcg cttggcctcc caaagtgtc ggattacagg tgtgagtcac cgcgcgcgcg 116160  
 tgcaagttag caagtgtgag atgcatggac ttgaggacag aagaggagcc cgaggtttct 116220  
 agtgggggct atgcacagtt ctgtgggtta aaggttgttc ttcctcttct cctgagctgg 116280  
 ggagctaggg caagctatgt cttatgtctc ccaagagtca aaacctttgg ctcttcttaa 116340  
 ataccagcct agctcatact aacactggat gaatacattt ccacaacaga aactcaaaat 116400  
 ccattctgtt aaggtatgat gacaaatttc atttagaaca acagtatgag agtcattatc 116460  
 aaatgtgaag caaggtgag cacaaggctc ttatttgttc ctttaagaaaa aaaaatttat 116520  
 ttcttagtga gtttaaatca tcaaccatgc ttcagggaaga aaaatgtagc ctggcctcct 116580  
 ttccattctt ttgtgattgt ggaaaagtgg accttaaga tttcactgaa agattacaga 116640  
 cctgctcatt aattttacat tccaaacctt tgactggtga taaaatgaac aaaagaattt 116700  
 aaaataaagg tggaaaaaaa tggatccctt ttcattccgt agggccctca ggctatatac 116760  
 aaacacaata cacaaaacct gcattcattt cctgggacaa tgttatgggc tactaagtcg 116820  
 atcttttag aagagcagtg agttaatgga ttgttttctt gacaatattt gtagtacagt 116880  
 aaaaaataaa atgggggaaa tttttttatc tggcttttaa atgttatgtt ttatggttgg 116940  
 gttaatgttg tggctgatca gctgttacat ggtgcatgaa tattgactgg attctccatt 117000  
 agggctgatg gggcatcacc gaccagctat gtgatctatg ttctccact ggctcttgc 117060  
 tggcatgaga gatacatttg caggatgagg agcaggaatg agagctatcg attgtagaaa 117120  
 cagaatatag tagactgtt tgacctgtt caaaattaa aaatctgtt ccatatcttg 117180  
 cattacactc ttttatttta tttcttact ctgaattcct tcatgaaatg cacaaactcc 117240  
 accctgtata tacatatata tatatgttg catttccctag cttttaaaaa tagtttaag 117300  
 tataccttct ccaaaggga aatactttga aatatttttg gaagagaaat tgtattagtt 117360  
 tagatccata ataaataatt atattgtttt aaaagctgggt gcaaagtaac attttcttat 117420  
 attgtagaac tctaagcctt ttggcttaga aatggagagc aatattcttc tgcttttcat 117480  
 cacttacagg gatattcttg gtattattat ataaatttta ttttcattaa atataccata 117540  
 aaattccaaa tgtgtgttgc gaattactgg cagaaaaatg ttggtaat tttccagtgt 117600  
 gacttaactg tcattataat taagtaattg tctccacca aatgtttgtt taggcaatac 117660  
 atttctgcatt atttgaaaac ctaataaaaa gttttattaa agctgttcac tcaaaaatag 117720  
 acaactgtgc cccgtgagc agtagttta aaccaaatca atgaatttgc tattatgtag 117780  
 acaagaaagt gtccatcat tgacattagt aatgtttcca ccaagttgaa cgtgatgagg 117840  
 gagctgttag ttgttacctt ggaaactatt caaaaggtct agagctgctg aaaaacagaa 117900  
 gactctgggc tcacccctgg gattccgctg cctttcagac catctaggat ccaaattttg 117960  
 ggaatgtatt ctgtatgtct gaacatgctt tactgccatt ttagactgat tgtatcta 118020  
 ttttaagtta tgaatgccacc tctcttttct cctattttc taaactttct ggaagacgtt 118080  
 ctgaaatcat tcaagtttat tctcattttt tatatctcag gactagtcac tcaatcttta 118140  
 aataaccact tttgatcaag agattactgt acccttctga tacaacttaa gatcataaaa 118200  
 acaagttgat ccaattcaat agctttttct tatattataa ttcattgcctt ggccgaataa 118260  
 ttccaacttt tctttttttt gaaatatatg ttaggttttc ctggttctca atcaatctac 118320  
 tgcagcactc attctttctg tattctgggt cctactacat cttttatcag tgtctctttt 118380  
 tgcattcatt gatcagtatc tcttttttta aatgcagaaa aaatagagtg tgcttataaa 118440  
 tgtttaacta ttagccagct atttaacttc ctaacactat ctgatatgct tgttattaaa 118500  
 tccaactcat tttctatgcc tttgatccgc cttatacact actaaggaca gaagagcaat 118560  
 ggagtcgatc taaaaccaga gaaccagaat ccttactcc cacagagaga ttgcctgtgg 118620  
 gctttaggca tcgggttact tgtctctaga gctcaagttc acaagcacca gtagcccat 118680  
 gggtagaagt tgaacctctg ttctatacag catgttgatg aatgcattac ttctgcctta 118740  
 gtcagtctgg agcacagtaa atgtcaacta gcagtgacgc accggaacac gacatgcaat 118800  
 ctgctcactt gactcatgct gctttcagaa gagaaggcac tgggttaaat ctgcatggca 118860  
 agcctgcccg agctagaatt cgaatgagaac cctcagggtta atagctttgc tctttgagag 118920  
 agctctaagc tcttttcatt ccagaccaat ttcacattta catactacct gtcactaaag 118980

```

ccccaaaggtc cttcgccaac aagaagcaga tatttactca ctaaagttaa tgtcccagcg 119040
ctttctagcc tttggacaca gtttttaggc aacaaggaaa ataatacata tgtataaatg 119100
gccattttgtg ttccctaaaat agcaaggggc tataaccocaa agcctagaat tctttttcgt 119160
tttttttttt gttgtgtgtg ttgtttgttt gttttgtttt gttttgaaac agtcttactt 119220
tgtgtgcccag actggagtg c aatgggtgcaa cctagggtca ctgcaaaactc agcctccctg 119280
gttcaagtga ttctcctgcc tcaccctcct gggaacctgc aattacaggc gccagtcacc 119340
atgcccgccc agtttttgtat ttttttagtag ggatgggggt tcaccctgtc gttcaagtgt 119400
gtctcgaact cctgacctca ggtgatccat cgtcttgccc tcccaaagtg ctaggattac 119460
agggtgtgagc caccacgccc ggocctagaa ttctttttct taaagacaat tataatggta 119520
attccgttta aaatgtcact ataaatctat ttcattttatt actgcatgat tttaatctat 119580
tattccactt aaaagaaaaa caaatcagag aaatcagcaa gggggcaata acaaagtgtg 119640
gttaggggtac acagccact ttgggacccat attcaatgaa gaataagatg agagaaagat 119700
aatgtgattt cttttattat tattattatt attatacttt aattttctggg atacgtgtgc 119760
agaatatgca ggtttgttac ataggtatata acgtgccgtg gtggtttgct gcacccatca 119820
tcctgtcatc tacattaagt atttctccta atgotatcct tgccctagtc cccaccccc 119880
cgacaggccc cagtgtgtga tgttccctc cctgtgtcct tgtgttctca ttgttcaact 119940
cccacctgtg agtgagaaca tgtggtgttt ggttttctgt tcctgtgtta gttgtgtgag 120000
aatgatggtt tccaacctgg atgaagcttg atttattaag accaatagta attacagagt 120060
gtgcggtcca tgagtgaagt aggatgtttt ctgtgacatg tgagagaaag cccaattcaa 120120
tcttccttaa gcataaaaag ataactttatt ggcattgatg agtaacagtt caggagttaga 120180
tcctcaagag tggcttatct aaggcctcaa agatcataat cagagtctaa ttttaactcat 120240
attatatgta ttgattaagc ttcattctatg taagctctat acttaagctc cagtccctac 120300
tcagaaaagag ggtgacactt cctgatggac ataactgaaa cccagaaatg ggtcacaaa 120360
agccccgagt tagactaact tgtgaaagag cccaccttag aaccaatcat gatgcccaca 120420
gacttggaag gtgctgattt tgaagcccca tcactgaagc agtgaaaagg aggcccgcca 120480
tgactggccc cattttgtct ctaagcccac ctttaggtaa acggcttttg cttatctctg 120540
catgtaggcc aagctggcta cgggtggaat ttggtttata gtttcacttt aaagcaaggt 120600
ttgtaataat tccttcccaa aactaaccac ggaggaaaca agaagggcgt acacataatt 120660
aagaacatta ggttaaaagt tataagagca ttgtgacctg accaaggatg aagaacttca 120720
gattccaact ccttccctct ttcctcttcc ccagatatat atgttccctaa aattctatta 120780
ggatggttct ttaggacact aatcttccat cttctcagtt tgcttgtcct ccaaaataaa 120900
gtcagcttcc ttgctgaac atcttgtctc ttgacttatt tgctgttggt tggcaagcag 120960
tatgggcttt gaactcagct acatgtctcc aagtctgtga aaatgatgtc ctcttagagt 121020
ccttctgttc acacctgtct ccactctccc gggctggaga tgtagtcagc ttgacctatg 121080
cagaggagat ggtctgttcc ctcaagccat atgtgaaatt cctgttgcca taagaagatt 121140
gatcctgggg ctaacaaaac atcaaagcct attatagata atatttaatt aaatctatgg 121200
atacttctag ttccggttagc ttccagccaa ttgtcaatgc cacttcataa aaattaagac 121260
aggagacat ttttgagca cttagtatag actagatctt atgttaaatt atttagacta 121320
ggcagacatt tctcatagca gttgtgtaag taattataat tgtgtcttct tagagctgag 121380
gaaacatata ccaaaaaatt ttaagtaata ttgacaaacc cacatagcca gtgagaggag 121440
aagcaggagc ttagaaacaa gottgtctga ctccagagac tgcagctttg cttataaaga 121500
cagaatagag aatgttgctt gaaaacatag gttcaaatat atgtttcacc acttagtagc 121560
cattcttctt tgaattctcc acctagattt gggatatttt cttctatagc taatgcctaa 121620
acacttaaaa gtattgttct ttttaaggaca ggcaccttca gagcctgtca gaagtgtcga 121680
tcaaatttct tcttagtcc tacaatttct attagcctct ttttctatgt cagtactct 121740
ggagtttact taagtcaatc aaccttcact tcaaaaactc ttccctggaa tctttttcct 121800
tttcttttca tttagcttag ctcaaagtct atgcagtctt ttttttcttt tttttgagac 121860
ggagtctcgc tctgtcacc aggctggagt gcagtggcgc gatctoggct cactgcaagc 121920
tccgcctccg gggttcacgc cgttctccta ttctcctgcc tcaacctccc gagaagctgg 121980
gactacaggc gaccaccacc acgctggct aattttttgt atttttagta gagacgggg 122040
ttcaccttgt tagccaggat ggctcgatc ttctgacctt ctgatccgcc cgcctcagcc 122100
tcctaaagca ggggtgtctt tatttattta tttatttgtt cattttattt tattgttttt 122160
tgagaaagag tcttgtctctg tcgcccaggc tggagtgcag tggcgtgatc tcagctcact 122220
gcaacctcca cctcccaggt tcaagcgatt ctcccacttc agcctccoga gtagctggga 122280
ttacagggtc ccgccaccat gtctggctaa tttttgtatt ttttagtagag atgaccttct 122340
atcactgtgg acatgtggc ctcacgacct cctgacctca ggtgatctgc ccacctggc 122400
ctcccaaagt gctgggatta caggcgtgag ccgctgcgtc cggccaaagc agggatgtct 122460
ttaatgccgc atcttcagtg cctaccatag agcccgcat gcaatagaag tgtgcgctaa 122520
ctcttccact tatccctggg ttaggcccaa ctctccagc tatagtagta gagtcaagga 122580
gatcccagaa aaatctgttt tctctgcacc caggctttta agtcatcatg agcaagactc 122640

```

tgccttttct	tagtttttcc	aggtggtgaa	gatgtggaac	aggaaaaacc	acataaaaaat	122700
gcaatcacat	gaataaatat	tgcctacaat	ggctggattc	aactgaagaa	cttacactct	122760
aaatcaggag	cacagactca	tgtttatgcc	tttgtcccag	gcaccaactt	ctcttgaaact	122820
tctttgctga	ccctgcctga	aagtggagtt	aactaaatct	tgtgtagatc	tgtctcatat	122880
cactatttag	cccatttgca	gataggtatg	atattgttaa	aatgatgaaa	atgccccacg	122940
ttggactcag	ttggatatcc	tgctccggca	gatcaggttc	taactggaat	cttggttcac	123000
ttcacttttg	taggctaaga	tattctattc	gtacttgact	gagcattgta	ttgatgatcc	123060
ctgataatgc	agctcttcaa	atgtaataaa	tttcttttgt	aattgactgg	ccatgcccac	123120
tcaagagata	tatgtttggt	gcctgattga	gtgtcctaagt	aaagaattaa	atgaataaat	123180
gaaagagtaa	cataagaaat	aaacaacact	gggacaatga	actgagactt	catttattaa	123240
gaagcaatta	ttgataaatt	tgtatatgat	tttagactat	aactaatata	ctccaaatgc	123300
aatagtagac	aagatagaaa	tggctccctc	cttcatactc	gtgggtgaaa	tacattgtat	123360
tgctgatggt	ccatagcttt	tttacacatg	ttgttttcat	ttgtttcaac	taacggtaag	123420
gagtggtgat	atttgaatcg	gccagtgcga	aatgcagtga	cttggaaga	gggaatcact	123480
gttagtgaat	cataaacaat	atttgaagg	ccttcagtgg	tttattaact	tcaaataata	123540
taactagatg	agtctcggtt	ttacaagtaa	ccaggaagta	atgcaagaat	cctagcgagg	123600
gtggtatcaa	aaacataaag	tttatgtata	aaaagcattt	gaactcaaga	atttgacata	123660
gaccttagtg	aaaatgcagc	tatttcttgt	cagtttaaaa	gaaataaata	atttggttaa	123720
ttcaaagcaa	gggactccca	atcaacttag	ctcacaggtg	atttcaaata	aacataatgc	123780
ttattaacag	ttccaatatt	tgactttttg	ctcaacctaa	attcagcaat	gcaatgtgct	123840
gagatatatc	ccagtaaaaga	gcagctgctt	ttgttttgtt	ttatttttgt	ttattttatt	123900
tcacaaacaa	gtttctgcat	gagcctgaat	ctttacaatg	tgaaagctgt	ttggattcag	123960
gtaggaaata	cgtttgaact	gatgcatacc	aaatttccaa	aagttcacct	ttttatctta	124020
ctgcttagct	gttgaataac	acaacttttg	cttagtcaag	ccttctttca	ttttcttgaa	124080
gatgaaaatg	taattcttga	gttacttctg	aagaacaggt	tggggaaatg	caaataagag	124140
cattactctt	ctagagtatg	aaaagattac	caaatttagg	ttttctggtc	agtaaaatta	124200
tattttccta	acttggttac	taattacaga	gtcattttaa	tagattatca	tttcttcttt	124260
ttccattgcc	atctttttta	atgatgatat	ttccagaact	gctctattta	actgttcagt	124320
gaaattacat	tcattgaagct	gacaattatt	cagaagtaag	tttcttatgt	aactctccca	124380
ttcttcatat	tctttgttct	taattttatat	ttgggtctga	aaattgataa	agttactagt	124440
ttagaaataa	ttactgtgta	cttcttattt	tcagaaacat	aaatgagggt	gcacttctcc	124500
gcagaacctt	tttcacattc	ttttaacaac	caaatcattc	ccaattttct	ggtgtgtaaa	124560
ttccatgtga	tcagtacaaa	taaacaaaact	gacaagattt	gggcagttgg	aacccattgt	124620
tttacaacgt	ctttgtttta	aatctttcat	aattcatctt	cagcaatggc	cttcctagca	124680
attagttacc	tttagtgtct	tcattttcgt	ttaaatgttc	aatgatcagg	ataattagaa	124740
ttaatttctc	attgtcattt	tatggtaatg	aagcttctgc	cacttatcac	aagcatcctg	124800
caggcacagt	gttaccacac	aaatatgtga	ctcctcactc	ttcacacact	ggtttatggt	124860
ccctgggttaa	agtggtaaca	gtaggacctg	gcctgattcc	agtttaaatg	tatgccatat	124920
tagatgaaag	attgtgggta	gtcatgctgg	tattcttgcc	ccttgtggat	ggcattctat	124980
gtgcaactca	gttctcctta	agcaacacaa	gagaattttc	gtctocatta	gtctgggtgt	125040
tttacagact	gggattagtt	ctacagttag	aacaaaaggga	aatatctttg	tgaatgaaga	125100
tggaaaggcaa	tgtcaattca	aaactttata	aaaacttcag	atagaattca	ggaaagcaaa	125160
gggttgaaaaa	cttagtacca	gaaaatctat	gaagaagaaa	gtgagtcagt	gaattctcaa	125220
gctggaagag	ccttcttgca	ttatctggtc	taggattctc	cagccacaaa	aaaaaaaaaa	125280
aagtaattgt	gtattatcat	ccccattcat	tccctgtcat	atctctgcct	aagcaattgc	125340
agtagaattc	atcctatcat	ttgtatcctc	ttccttactc	ctgggaaaca	tttagatagg	125400
gtaaaggatt	tggttgagca	gaaacctggg	gtgggaaata	gggagcatat	ctagggttgt	125460
caaagaagggt	atttaaggct	gccctgtgaa	atcataccag	tttcaaaatt	agaaacttga	125520
tctttttttac	agctgacttt	tcttcaccac	attttaaact	tgttctttat	tctgcatttt	125580
ctcaagaagg	gatatgatat	taccaagaga	aatacttgct	ctcaataacc	ccaacctcat	125640
tttcttatga	gtttgggttaa	gttagaaaaga	aggtcaaact	tagtagctta	aagaaaaaaa	125700
cacattttct	tttgcttata	attgttgagg	aaggaaattg	gaaagggcta	ggcagggcag	125760
aattggaagt	tttctgtgtg	gtctggcaga	tgttggttag	tgctgcaatt	atgtggagtc	125820
tatttaaaact	ggccatccta	tggtctcttc	atgtggctgg	tggtgaatgg	tggtctgttg	125880
ctgggaactc	agccggggat	gacaagtgga	tcacctccat	gtgtctttgt	agcacagcag	125940
tttcaggctt	ttagaatttc	ttccagtggc	tgcatataac	ctgaaagaat	gtcccaagag	126000
agtctaaagg	agaaatgtgt	ggcttattta	ttaactagtt	tggaagttac	aatgagacat	126060
ttctgctgca	gtttactgat	cacaagagag	ttactaagat	tgatcgagtc	tctacccatg	126120
tgaggggaggt	gaatcaaaga	acttggttga	cagggttttag	acctgtcata	gagtgtaaaa	126180
aagctaagag	gcttcatatt	caagggaaga	aatactttta	agttctccag	ggatgaggct	126240
agtagtgatg	attcagggca	agaccagtga	agtaagagag	gctgggacat	gcattgaaga	126300

agctcactga	gaaagtcttc	tggtaatggg	ctgcaatgtg	ctggagttag	tccacaatgg	126360
cttgcaagat	ccagttactc	acttggtttc	tcaattccag	ctctgagttc	agtgatgtca	126420
ttgatagttt	gaaaatgact	acagtggaa	atttacacca	cattaattgg	cagactctac	126480
aaataaggg	gttttttccc	caggagacc	agttgctgaa	tagttaccaa	ttggtaattg	126540
ggcaacttag	aaactacttt	taggagaact	ggactgtctc	atagcaattc	tatgtttgtga	126600
actttgggg	taaaaccact	ctgcttttgt	aaccttttca	gtaaaatttg	cgagggtccc	126660
tattgctttt	gggtgaatga	cattttgggg	gataatctaa	tgtgattctg	aatttgagtc	126720
aacactggct	gtttcagaac	agcacatgtt	gaagcaggag	cctgatgcag	aatacgtcat	126780
tagagggaaa	gtgctggtaa	ttagcagaaa	ttacaggaa	ctctgtactt	caaagacttt	126840
aaaattgagt	gtgtgaacca	agcacatttg	agtttccaga	gaaggaatga	agctccagtt	126900
gacactttgg	ttatacaact	gcattgataat	ttatcattta	cagtataact	ttatgtacct	126960
cataatgttt	aaatgggtta	tagatcaaat	gcaaataatt	aaatattttg	catgacattg	127020
gacaagtca	ttaaattttc	tgatactcag	attcttctatc	tctaggaatt	agctgaaagt	127080
atcttttatg	agagttgttt	cagtgaagtc	tgtaagatta	aaaagaaaaa	taagagacca	127140
tcattagcca	tgagtttagt	aaatatttcc	cagtgggtga	acaccaattc	caagtttatg	127200
taagttgggt	ttgatgggaa	ataagcaatg	attaaacaaa	ggaaaatcag	gagaaagatt	127260
gttaacaaag	ggcaatgctg	tataagtga	aaaaaaaagg	ggaaaaccac	agttttttga	127320
tttcacaagg	aagcagagg	ctggctttca	ttcatgcccc	actccagcct	cagttccatt	127380
gcttcttgcc	ggcaaagctt	tctcgtctctg	acagaaggca	agcaaacccc	aagaggcccc	127440
aaggtggcct	gtcagagggg	gtaaaatata	tagtttttga	attgtataat	attaatgttc	127500
ttatcagagt	gagaatgtga	gtagaaagca	gagattttta	taatatgacc	caagggaatc	127560
aattttgaac	attttacagt	attaacaaaa	atgttttagaa	ctttagaata	tatttaactc	127620
aataaaaaat	aaatatattt	tatgttatgt	atgaataaca	taaacaaatg	tttgatttag	127680
ttcagagttg	atattttaaa	ataataaatg	ttttgttcta	gtccatttaa	ttatgggtct	127740
tgattggcaa	ttctggaag	gactcagaaa	gtaaaaaagg	agagtaggga	aagcctcaag	127800
cttagatacc	acctaagtca	tcattaatag	ctattggtag	aaatatgggt	ggtgaagacc	127860
attctgatgt	gttcccagat	cctactgggg	agaatgggat	tggaaaatgg	agaaaagtct	127920
atctttgtta	tgaagtgtca	cataacttgg	ctgagttgtg	ttcctgtcct	gatatttttg	127980
ggaaggtaga	acttgtgagt	gatgaaattg	tatatgtggc	tgcagcactt	tctaaacaac	128040
actttgaaag	agtagcctcc	ttcctctttg	ctgctcctag	taaaatggga	gaagagaaaa	128100
attattagag	gatggaattg	ttaatcaaaa	ggaaagcaaa	acttaaatat	ctggaaaatt	128160
cttagtctat	ccatattgta	aaaaatgagg	aagtatcttt	gcaagagaag	accaagggtg	128220
tgaccaaagg	acaatctaac	agaattactc	agtcatttaa	cagaatctag	aacctgttgt	128280
cnagaccat	ggaaggataa	tcctaaaagc	aattcaaaaa	ctatcagaac	tgccaccctc	128340
atcgacagac	caaatgcaa	aagtcggggc	agcagggtgg	tttccaattt	aaagaaggga	128400
ccaccacgac	cccgtgctgc	ttcacattgt	ggactctacct	cccttactct	gttaccatgc	128460
accttggtag	ccccagattc	agttctgggtg	aactctggca	gggaattgtt	aggggggctg	128520
agccaccaca	gtgaacttct	aataggccaa	tgcccagcag	agccacaaga	gtagggtctg	128580
ccctgagacc	tcagacctgg	agagccacgg	gcataaaatt	cttggtctga	agatccacag	128640
gcacacatct	ccaatccatg	agagctgcat	tgtaagtgtc	atctggcaaa	gtcatgtagc	128700
caggggttcc	cagagccttg	gaggtccaaa	ctctgggggg	caaagctaag	ggagtaccac	128760
ccaggggtgct	cagaaggac	ttttcccaa	atgtgggttg	aagggcagag	catcaaaccg	128820
aagaagattg	ttattgagac	ttatgtctata	gtgttttctc	tgttgggctg	tggacttagg	128880
gcctctttcc	cctttcctct	ttcctgtttc	tatgttttag	aatggggatg	catgttctgt	128940
gcctgtccca	tcattgtatt	ttggaagcac	ataacttgtt	tgatttcata	ggttcacagc	129000
tagggaacaa	tttgctcaa	aatgaatctt	acottgagtc	tcattcatac	ctaacttaga	129060
tgatatttag	atgagatttt	ggactttgaa	ctattgagtt	gatgttgga	caagtttaaga	129120
cttttggggc	tactgggatg	gaacgaatgt	cottgtctgt	gagaaagaca	tgaattttga	129180
gggctaggg	cagaatgctg	tgggtctgaat	gttcatgtcc	cctcaaaaata	tgtataaaaa	129240
tttgattttc	caaagtgata	gtattaagtg	gtggggcttt	tagaaggtga	ttatgggtga	129300
gcctcatga	atgggattac	tgcaattaca	aaagaagccc	aaggaaaacc	cctttctcct	129360
tccaacatgg	gaagatgtag	gtgaaatatg	acatgctttt	ctgtgaacct	aggttaaggat	129420
ccaccaccaga	ctccaaattt	cctgggtgct	tgatcgcaaa	ccttttagcc	tccagaacta	129480
taagaaataa	atttctgttt	ttaatgaagc	tgtgcaattt	atgatacttg	gttatagcag	129540
ccagaagagt	ccaagacgtg	gtcctacaat	gaagttagag	aattttttta	agtgtctgtt	129600
tattgttaca	aatggcaaac	aaaaggtcat	tatgcttaca	cacacaataa	atgtgtacaa	129660
tgtgtatgag	taagtttaaa	acaacaattt	tttttaatat	gaaagtccca	gttagtttta	129720
aaagaaaatg	aacaactgag	agtaataaaa	cotttgttct	tgttttgaca	gtgatgggga	129780
agggttttca	aaaataatga	gttggcttca	gtttggaatt	agggaaattac	ctaccataaa	129840
actgtctttt	agatatccta	aaatatcatc	tgttttgaga	tacacaacat	ctaaaggcca	129900
atgactttat	tttctctcac	agagaaagga	gacctaaaga	ggattatttt	aggtttggat	129960

tcagaaatgg	ccattgtgta	aagtatgtga	gagcagattt	gaatatttgc	catagtgcag	130020
aaataattgc	ttatagtacc	tgattttttt	aagttattat	ttcatccata	atatatcttt	130080
ttcttatatc	tttaaaattt	ttacaagata	aaatgaaaat	tgaactaact	tagacaaagg	130140
ttattttcaa	attataaaag	tatatagatt	tcaaaatggt	tattgaatac	ttatgcacaa	130200
tggtcaaaac	acttatattt	ttatataata	cactatgatt	tttagcattc	tgagcaaaata	130260
tgtgaaacct	tttagatatt	tgcatatttg	ttacaaaaga	agtttagtgaa	acatgtcaca	130320
tatatacagt	ttttttatat	tagcttgaat	ttcttaaaat	cagcctttac	caggatatatt	130380
agtcattctc	cactctgcta	ataaagactt	acctgagact	gggtaattta	taaagaaaag	130440
aggtttaatt	gactcacagt	tcagcgtggc	ttgggaggcc	tcaggaaaact	tacatcatgg	130500
tggaagggga	aacaaacacg	tcgttcttca	catggcaaca	ggaaggagaa	gtgctgagca	130560
aagaggggaa	agcccttgt	aataccatca	gatcttatga	gaactcattt	gctatcatga	130620
aaacagcagc	atgagggtaa	ccagcccat	gattcaatta	cttcccaccg	ggtccctccc	130680
acaacacatg	gggataatag	gaattacact	caagatgaga	tgtgggtgtg	gacacgcaa	130740
actatatcat	ttcatccctg	ccctctttca	agtcctcatg	cctcacattt	caaaacgcaa	130800
tcatgccctc	ccatcagtc	ctcaaagtct	taactcattc	cagcattaac	tacaaagtcc	130860
aagtccaaag	tctcatctga	gacaaggcaa	taccctttca	gctatgggcc	tgtaaaatca	130920
aaagcgagtt	agttacttac	tagatacaat	gagggtagag	gcaatgggta	aatacatcca	130980
tatcaattgg	gagaaattgg	caaaaacaga	gaggccacag	gccccatgcc	agtcacaaat	131040
ccaatagggc	agtcattaaa	ccttaaaagt	ccaaaaatgat	ctttgactcc	acgttttcacg	131100
tccaggtcat	attggtgcaa	gaggtgggct	cccacagcct	tgggcagctc	cacccctgtg	131160
gctttgcagg	gtacagtcct	cctcccgcct	gctttcatag	gctggcattg	agttcctgca	131220
gcctttccag	atgcacggtg	caagctgttt	ggtggatcta	ccattctggg	atctgaagga	131280
gggtggccct	cttctcaaaag	cttcaatagg	cagtgcacca	gtggggactc	tgtgtggggg	131340
ctctgacccc	acatttcctt	tccacactgc	ccatgcagag	ttcttcatga	gggtcccatg	131400
cctgtagcaa	acttctccct	ggatatccca	gcatttccat	acatcctctg	aaacctagcg	131460
agaggttccg	aaacctccaa	tcttgtcttc	tgtgcatcca	ctggcccaac	accacgtgga	131520
agctgccaag	gcttggggct	tgcactgtct	gaagcaacag	actgagctgt	accttggccc	131580
cgttttagtca	cagctggcat	tgagccagtt	gggactcagg	gcaccatgtc	ttgaagctgc	131640
acagagcaag	ggggccctgg	ttttggttca	ataaggcatt	ttttcatttt	aggccttcag	131700
gcttgtgatg	ggagcggctg	ccatgaaggt	ctctgacatg	tccctggagac	attttcccca	131760
ttttcttggg	gattcacatt	tggcttcttg	ttacttatgc	aaattcctgc	agtgtgcttg	131820
aatttctctc	cagaaaatgg	gtttttcttt	tttattgctg	cttcaggctg	caatttttta	131880
aacttttatg	ctctgattac	tcttgaatgc	cttgttgctt	agaaatttat	tctgccagat	131940
accctaaate	atctctctca	ggttcaaagt	tccacagatc	tctaggtcag	gggcaaaaag	132000
ctgcagctct	ccttctgtaa	gcatagcaag	agtcaccttt	gctccagttc	ccaacaagt	132060
cctcatcttc	atctgagatg	accacttcac	tgtccatata	actatcagca	ttttgggtcaa	132120
agccattcaa	caagtttcta	ggaagttcca	aattttccca	catcttccca	tcttctgagc	132180
cctccaagtc	tctagaaagt	tccaaacttt	tccaagtttt	cttgtcttct	tctgagccct	132240
ccgaattggt	ccaacctgtt	cttgttacct	agttccaaag	tggtttccat	atttttgggt	132300
atcctttatg	ccccactccc	ttggtactaa	tttaactgcat	cagtttggtc	tcacactgct	132360
aataaagacg	caccccaaac	tgggtaattt	ataaaggaaa	gaggtttaat	tgaatcacag	132420
tcctgaggct	gggaatgtct	caggaagctt	acattcatgg	cagatgggga	agcaaacacg	132480
tccttggtca	catggcagca	agaaggagaa	atactgaata	aagagagaaa	agcccttac	132540
aaaaccatca	gatctcacga	ggactcactc	actatcataa	gaacagcctg	aggaaaccac	132600
cccatgattt	aattacctgc	taccgggttt	ctcccttgac	acgtggggat	tattggaact	132660
acaattcaag	atgagatttg	ggtggggaca	tggccaaacc	atatcaccaa	gatataaaat	132720
taagttgcct	cactttaata	acaagggttg	gttacatttt	aaaattatta	ttaaagtata	132780
tatgtgtgtg	tgtgtgtgtg	tgtgtgtata	tatatataca	tatatgtctt	taaagttatg	132840
ttaagtgttc	catgtactac	tgtattgagc	acatagaggt	gctcagtaaa	tgtcaatgtt	132900
aggtgaataa	atgagcggat	aaaatatgga	ccaacttagc	tggctcagtga	gtttatgggt	132960
acttatcttt	gagaagcatc	ttttgccttt	ctccatattg	agattgttta	attgagctgt	133020
gtaacttttg	tatatgttat	tatatgtgtt	taaaattttg	atataatttg	tctaccttag	133080
tagcaaattt	tattgagaat	atttttttct	caaaatagca	tcccttttaa	ctagaaaaaa	133140
aaacagtttt	tatagctctc	ttttgaacta	aataactaaag	aataaattgg	tcaagttctt	133200
tctctgatca	tgaagtacct	gcaaactgct	tcccataaat	ggcagaaaaag	gttaagagac	133260
attaccagga	tggtgcctcc	aaattgttgc	tggtatcatt	aaaaatgtca	acaaaaatgt	133320
ggagcctaca	gccttgtaaa	aataggaaca	tttaaatcca	tacttttagtc	atttctctc	133380
cctgatttat	gttttcttgg	ttgtttgcag	gcgcagtcac	cagcagtgga	agagctggct	133440
catgtgtggc	tcaggctgac	tccacaatca	gagtgattat	gtggaaatca	atcagaattt	133500
tcttcttaga	agatgtttgt	tctgacataa	tgaactgacat	tccagctaat	ttttattcat	133560
aggtctctcc	ttgcacaatg	atctctattt	ctctcaatgt	tattctgtgt	ctttcttgc	133620

tgctgcaca	ttctgaaaaa	aaaaggaaa	caattcccta	cattttat	tgatacttaa	133680
tgttgtttg	taacatattc	agaacacaat	gcaactttat	atattagagc	acttataata	133740
tattattagga	aaacatcaat	taggtaccac	cagttatttt	agacaggtat	ctcttaggta	133800
aaatgtggtg	aaaatagagg	ggcaatgata	gaattcatat	gtgaaagagt	caattggcca	133860
tcattatgcc	aaaattat	tgtggttcag	gggacatttt	aaactttaat	ttttaagaat	133920
taaactttta	ttaaatttat	tgtagtttaa	atgcaatcac	aaggaaagac	tctgaactgg	133980
aaacaccaag	ccaagctctt	tcctagttct	ggtcccttat	aaaccctgaa	agttaatgca	134040
attattgtaa	tttttaaacc	accgagttct	cagggtgatt	agttatgcag	caatagatag	134100
cctgaaaagg	accattcttt	ctggcactga	aaattaaaga	gagtgattaa	caacataaga	134160
atgtagtttg	tgctgtctgc	cacagattta	tgttgtagct	tcagtgattg	cttcaaagca	134220
cccttactaa	ttgccagcct	attgctgata	cctagctaaa	ttagcatgtg	ttagttttga	134280
taccactata	ggcattttaa	ggatagctga	tacattctat	ataaaatata	attattttcc	134340
tatctttagt	ttatcctaaa	gtagcagcta	ttatattcgc	ttcttgatta	gctttataca	134400
taaccagaga	aagtaaaata	tcacagagta	tatttgaga	cataaaagtc	tttggaacaa	134460
gctcaagtaa	tgaattcagga	tgctattaca	tcctccacat	gacagtaaat	atattttata	134520
ttcctaaatt	ttaagctttt	aacaaaagca	aattttat	tggactat	caattttctag	134580
ggaaaggaaa	ttatcattta	ctgtctttga	ggctataatc	tatttagcca	ttatttttat	134640
attgatcaat	ttataataaa	ggtaagagaa	aatactacca	catatatgtt	gtttatataa	134700
ttttaacttt	cataatttaa	atactgggtt	tactgcaaga	tgtaaatatt	tataatagct	134760
tatatgttga	aacaattata	ctttatat	tcctaattcca	ttgccttggtg	tgctcattgaa	134820
gaaattgaag	ttcacaataa	tgactgaaca	aatgctcaaa	gctgggtctgt	ggcagaacca	134880
aaaatttcca	tgccaccttt	ctctttccat	attgggggtga	tctgccta	ttgactttct	134940
gcttatatgt	ctagactcta	cccctcattc	cttctgctag	tgtgtgtgtg	aaccctttct	135000
attaacccac	tctcgatttt	tctgttgac	aatcaatctg	tagcctgaat	gttgagtccc	135060
aaatgcatga	gctacctggc	cgagtaatct	aactagatta	gocctgagtat	atcactttaa	135120
ggccaatgta	aagttatgtt	gtaactacct	catagttttg	cagaggcaag	aaatagaagt	135180
taagggaagag	aagaagccaa	tcttgaatca	gttagaattt	ccttggaacat	agaaaatctc	135240
tacaagcaaaa	aacccatttg	attgaacaag	caagcttatg	cagttaaata	tatcaaagat	135300
cttgtcacia	aaagtgtaaa	gggaaagtaa	aataaggagg	gagcatgaag	aagccagctt	135360
atattaaaga	tccaagcag	ggatataagg	aaattcaagc	caaaaataag	gaagaattct	135420
tataaatgcc	ccaagatga	attcgatgtt	tggtcgatat	tccaacagat	ttagtaattgc	135480
tattttttaa	gtcaacaggc	tgcccatttt	tgccctcacac	tctgactgca	tgtaatctgt	135540
caagtactct	gatgagttct	ggagatgcta	atgcaaaaaa	gatataaaat	ccttaccttg	135600
aaaatgtagt	tcacaattta	gccatggagt	tatggcaagc	atcttgtctt	caatttttagc	135660
ccctcctcca	tcctctctcc	agacacccat	tgtcctcttt	ctcaaccacat	ccacccctt	135720
ttttgtctcg	cttttggtgc	actattttata	tgctctctca	tgctctaccac	aactattttg	135780
ggtagtaaga	acatcagtat	ggagacccct	taatcctaac	taatatcaac	tctatctgtt	135840
tggtattaat	tcttgtttat	tgaacttgag	gcattgagta	ctacaaaaat	tataatacac	135900
tactggtgaa	cattacttta	ttaatgatta	ggttccactt	ggatgtttgt	cattcgtagt	135960
tagctgttct	ctccttgtag	cctgtaaact	ttctaattgat	gtcacaaaaa	gagttttaag	136020
gatttggctt	aaattaaaat	taatacatca	agaaaatact	atgcctcaaa	caaaaaataa	136080
tcagtttaatt	tagtttttag	aaatctttta	caaaaaacgt	gataattta	aatctttatg	136140
caaggatttt	tgggatatta	aaattgtcac	agatttaata	atacgtttta	ttgtataaaa	136200
tttttctttt	ctcttcaaag	tggtgggaga	gtgaattctc	taagtaaaaa	ctgaagagta	136260
agatgttate	attaaacatt	tcagaaaatg	cacactatcg	tagactgccc	actgtgaaaa	136320
tgatcagcta	gtagaggttt	tggtgaacat	atagtaactt	ttatcaacaa	gtaatgtgaa	136380
tatgtgaaat	ctattgcttt	tccaaaagaa	agaaaacttt	acacttcttg	gtaagataa	136440
actttcatgt	aagaaaaaat	aaaaacataa	ttttagcaat	ccaaaccata	ctaaaatcat	136500
aaacacttga	ccacatgttt	gagtacccca	aaaattgttc	ctgtgtggat	tcttcttcta	136560
ggatgatcag	tcctcatggc	cttactactc	tccatttctg	tcccactgag	gttagccaaa	136620
gttcacttgc	cttgtcagtt	aagtttctaa	atcagatgag	ttatgttttc	tataattgag	136680
ttaaaccacag	ttaattgctt	aatagactct	gtctacacta	actgaaaaaa	aagaattaat	136740
tgcttgagga	aaatgcatta	tatattttat	tcacatctac	agtcacocag	gccaaagtga	136800
caaaggcact	ttcaacagtg	gccagggcca	atggcgacag	cggggccagt	gcttagctca	136860
gccatgtgtt	caatctagta	ctttgggatg	tggaataaaa	gaaaatgagt	tataaatgaa	136920
cttctcctta	gagcttaatt	cattattatt	attattatta	ttattatttc	tttacataag	136980
aaaataattt	taggagtgtt	ttggcaaaa	tttaggtaca	atttaagaaa	atgccatttg	137040
taaaaaaaaa	aaaaagacta	gnnttttttt	tttaattagt	ttttactttt	aatgattttg	137100
atttaggttt	taaaaatgta	tttctagacc	aaggtctaac	accaggtatc	tgagattttt	137160
ctaaattttc	tctgaaaaca	tgagacgcac	gtacagttat	ttaaaatttg	ctgcagccat	137220
ttttaataga	ttttttatgt	ttgaagagat	gccaatcgct	tattatagta	agctatgata	137280



gcaatgccaa tgaatcctca attccattgc ttaaataataa gtgaaataat aaaggaaaag 137340  
accagtattg aatatattgc ctcttcctgc tacactctaa catattcact acaataaatg 137400  
gaaagttgca catacaataa aaacttaggc aattttattta tttacttatt tttttattat 137460  
ttttttgata tggagtctcg ctctgtcgcc caggctggag tgcagtggg cgatctcgcc 137520  
tcaactgcaag ctccgcctcc cgggttcacg ccattctcct gcctcagcct ccctagaagc 137580  
tgggactaca ggcgccccgc accacgcccc gctaattttt tgtatttcta gtagagacgg 137640  
ggtttcaccc tgttaaccag gatggctcgc atctcctgac ctccctgacc acccacctcg 137700  
gcctcccaaa gtgctgggat tacaggcgtg agccactgca cccggccggg aattttattta 137760  
agtataattg ataatacagg caataatcaga aaatcttgaa taatttgtac tctaaaatga 137820  
atttatgctt ttcagtgtcg gaagcattat agaggttggt tcaactatga aatcattttg 137880  
tcaaacgttc tttgaagtct atgaccttaa gttaaagcatt taaaacccaa tgggttagaa 137940  
acgtccagag aaacaagtat attggcaatt gtaaataatt tttgaagtat ttagcaaaac 138000  
agcgaagtac aataagcatc ctattcacaa ttccaatgta aaaagagatc tggctatagg 138060  
cagaattata ctaagaaaaa ccctatgaag aaaataaagc ttgtttgtta cctccaagat 138120  
ttcaccatct ttttcccggt tccctgcac tatgccctac tgtttctgct tcctcattaa 138180  
tcttaaatcg acttaagtgg atactgggtt tttaaatatt tcttatctgt atcacaagtt 138240  
caacaaggaa aatatcttga taataaaaaa agccaaagga agcttttaa aatttaccca 138300  
tggattgtaa ttacccatga tccccatcct attcctttag aactgttaga taagagtttt 138360  
ctacacttga aactttgagt gtttctaaaa atatttcttg ctttgagctt accaagtga 138420  
tagataccgt ccagacaggg tgaactgaag caagagtccc tggctcgaaa gtgtgcctgg 138480  
catgttctaa taaaacaaaa aggcgcacg ggtgctccta aaagggaata aagtcagcaa 138540  
gtggggagag gggcatgtat gacttcttgt tgattattag aagaacattg agaaatgcaa 138600  
aaccactttg tggttttgtg aatccactgc tgttttcagg atacactgag aggggtaatg 138660  
gcacaaaagg gaaatctcct ggttgatagt ttgtaaatct agacagaagg tgggtttggc 138720  
tagaatcaag gtaatagcaa gtogagttat cagaaatgtg gaaattgggt ttatattgag 138780  
aaacaagccg atggtgctgc tgagaattgc acgtgtagtg agggagaaca aagggactca 138840  
agagtatatt tttttaacca aagcagctag aaagatgaca ttgcctttgt acccttcatt 138900  
ctctaattgt attccccatc tctgccttca ttctctgatt acctttccca acaggtagct 138960  
ttagcttgtc cacacctaac tatgcttctg tctagggcct tgacacttgc cgattccttc 139020  
cacctggaat cagcttgccc aagatatcct catcaattat ctctcatatt tttcaagttt 139080  
ttatttgaat gtcacctttt caatacaact tgccttgccc cccatcttag attgtatccc 139140  
cctaaataca cttcctgtct accttctccg attcgttttt ctccatagcg cttatcccaa 139200  
tctaacaaac atattttact tattttattt attgcctatt tccccagtaa aaaggcagct 139260  
caataacaag gatgtctgct tgttttgctc actactctct ctccaaagcc tggaatagta 139320  
tttggcatat aatagcacia taaatatggg tttagatgat gaattaatgc agggatttag 139380  
ccaagctatt tactgtactc tgggtctctga gaaaatttgc aagcaggaga aagattgctc 139440  
tatggaaaaa cgtatgaatc aaagcaaaaa actccatcat ggtgtatcca gatgaaagt 139500  
gtttatttga atttagtttg tgataaagtc ttcatcataa ggaggatgaa aacaaaagcc 139560  
ctagttagta cagatcctca gagaagagct aataagctga gctccttaac atgactcatg 139620  
ctttctctaa ctgggggaa tggttggccc caactctcgc tgagtctgca cactgaggaa 139680  
aactcctcaa aatctagcag cctggaagga tgaatgcaaa agaacagtaa caccaattat 139740  
ccctctaccc agctgacatt tctctcccat ccttcatttc cttaccataa ccctctgtgc 139800  
tagcccatg ttcatggata atgtccaaga atgtacaatg atatacaaga tcatgataat 139860  
tgaattccaa cttctgtcat tttcttctgt gcccccctat atccagtgt tctgaactgc 139920  
tacagttctg taaacaaaa atcccttttc cccaatttta taactaaaag cttttacttt 139980  
tttattcttc ctagaatagc ctttcttato tcatatcttt gtaaatttta attcatcctt 140040  
taataatcat ttatagcagt tcttcattag tgaaattcct gatacatttc ccaccctctc 140100  
tgaaaaatta attgctccct tattgatttc ccattgattc taatagtatt gaaatatttt 140160  
attacaatta ctgggttcaca catctgaatt ctgcttggag ctgtgaagaa tctgaaggca 140220  
agaaacatat tctactttct ttgtataaaa tggacttatc acataataat ggctcattag 140280  
ggtgaatgaa tgattaatat taacaaagaa aattgtcaaa attaggtaag cttcttttaa 140340  
aaagttgaag ttaaacatta gccatatgtt ttctggaaaa ctgagcactt tctgatgtgt 140400  
ttgaaatcac attgagaaag tctgtctctt tgctactttt gtttactgga atgaagggt 140460  
actggaaaga tttcccagag aaaacaagac cacaatgtt ttgaagttag gaaaaatatt 140520  
ttaattgaaa tagaggtggg acaaagacaa gaaatatgac atgtgatata gtagatataa 140580  
gcgtatatga aatatatggg atgaacatcc atacaggttt cacaggactg aaagggtgcaa 140640  
ctggtcacaga gagtctggt ttgaaagtat ttaaaactca caatcaaaagc tctgatagta 140700  
ggatgcaatg tcaatgatta ctaatcgaga caaatttgc caagatcatt gtataattac 140760  
acccaaattc caatttgtga atagatttac tcatcgtgac ctagtgtata acatggaaag 140820  
ataaacaagg cagatacaca gaacattttg cagactatgg tggctgcaag aatagctttt 140880  
cgcttgattc aggccttaag tgttagcctg aacactcaga acgcgattta ttgttggaag 140940



```

gggattaacg aagacattta tttcagaggg tactttgttg ttgttgacat ttttgttggt 141000
gttgactttt ttgttggtgt tcttccctct catctgggat atttcccttg cccttctgac 141060
ggcttgtcaa tggggagggt ttcgttccctg ctttatctat tgcatactct tctgcttaat 141120
atgaatctat tacacagaat aaggaaatac atttagggct gactaatgta aaaatactct 141180
ttgatgagca gtacactgca acatagctca gagttttcaa atttattata ttggctcttc 141240
caaatatgtt attcatattt gctgagaggt gcttttataa atttctgctt tggagcatac 141300
agctcagtggt aaagatatgt gtttatctga attagctggc tctaatacca aatgaatatt 141360
tgtacaaatc tgtttttaac attgatatta atttattaga caattagaat atggcaaattg 141420
atgcaaagtt tggaaactgt ttcacatttt tttaatgggt tgtactcact cccctccccc 141480
atatacgaga ttttaaaagt ctatagggat atcttagtac atatttatgc agattacttt 141540
tttatttatg gaaggggatt gcttgatcct ccagggtggt ataaaactgc tatttcattt 141600
tgacatgata taaattgggt tagagaaaac aaattattct aaagtcttta gaaacaaaaa 141660
tgatgctggt ggagcatagt aataacgtaa ataaatctag aagatggaga tttttatatt 141720
tctagaacta gatggggttg tectggatca tgccctcgaa tttagtctct gatttttate 141780
caggggtgat ttttgacttg tataaaaacca tggaaaccaa gagccatcat aaagcaagaa 141840
tcgggtttaca atcttttgaa atgtgtattt acgtagacca tgaaataaaa aatattttgt 141900
tcaagaaaag ttgtttttaa tgctgctctg aacagcagac caaaaataac atttttaaga 141960
gcaagcaagt cgtttttgtg ttttacttgc ctgcccataa aatatggaat gttggacca 142020
gaagaatacc tgagtttagg accatacatt cttttaggta aaattttatt gaaaattaa 142080
gtatatatag aaaagtgcac aaatcagaag cctacggctc aatggatttc aagtgaatat 142140
agccatgtaa cttagcatgc aataaaaataa cacattgata gagcattacc cgaccgggc 142200
cttctcatgc ctcatcccaa aggtaatgct tattctgctt ttaattatt acatattagt 142260
ttggcctatt ctctgaaca caactcttca atgcttgtaa gttccatttt agaagatctt 142320
gcaagtggaa tttgccttct cccactagtc agtctttaac caggagatat caagatggaa 142380
tcagaggaat tctaagcact ttaagtgggt ttgcccgtgac agcatccttt gtgattactg 142440
aaagaaagct ggaagttgat aaagtgatcc atgttgaaag tacagattgt gctttgatag 142500
tggccatggt gtctctgcca caatgtttcc ctgcactatt tttgccaat attggattta 142560
acagtaaaag ttttaagtaat gccatattcc actaacatat aacaaaaaaa tgggaattaga 142620
aataggtagc aaatttttaa gatcagagaa aaaggaaatt tgacatttaa gaattacatg 142680
tggttagaga tctttatata tcaacatatt gaaagaaatt aattcagtc gttcaggcat 142740
gacgtttatt ttttaacaatt agtattctat tctctataat catgttccac agtttaaatg 142800
gcttaatgag ttcttactgg aaggggacta tttcgggaga gtataatggc ctttaatac 142860
ttcactcaag ctatttagcag gaaagattgt caggaaatat tgttactttt gttataaaat 142920
caattgaact ttactttgga tgaacaaga agtttgtcaa tatggtatat aacctccaag 142980
gtctttctcc cttttcctta tgctatttcc cctcgtctt cgccatgtgt tattgattca 143040
ttgaataat ctatgattat ttctcttctt aatagcctga gaagaaaaat agcttctaac 143100
taagcagcta ccaagtctt aagctagatg ctatgtttat attatctcaa ttaatccac 143160
tcagtgtagt cgatggtagt accctgttac aattgtgcaa aatggagttt caggaagata 143220
cacaatttgc tcattatcaa actctcataa catccacaga aaataattga aacacaaagc 143280
catatgactc ttaacctcta ctgaaagatg cttttatata gaatctaaaa tgttatgtta 143340
gatataatat attggctggt gaagttttat cttttccaaa acagaacaat atttttagtgt 143400
ttctcagtg aaaatagtga acactattca gttcagtg gttctcaag atgatattct tgattatcct 143460
catgagagag tcagtgtagg tagttcagtg ggtctcaag catagtaoct ggacaagcag 143520
caacagtgac atctgtgagt ttgttacaca cacacattca tgagccctac cttaaatctc 143580
tgaatatcca agcatgcat ccaggaaatt ctattttaat aggttctcga gatgattctt 143640
atatgcatta aggtttggga accactgaca tatttcttaa aagcacgaat tgcggaaaca 143700
gactgtgtaa gtttatattc atgttctgtg aaatactagc tatgtgacat tgaacaatta 143760
atttatcttt ctggtatcca gtcttcttca tttatgaaac ttatacaata atcatattctg 143820
catcacgttg ttttcagaat tggtgaacac acatagaaaa gtgtcatgct agggccaggc 143880
acggtggtct acgctgttaa ttccagcaat ttgggaggcc aaggcgggtg catcacctga 143940
ggtcaggagc tctagagcag cctggccaac acggtgaaac tccatctcta gtaaaaaaac 144000
aaaaaattag cctggcattg tggcagcagc ctgtaatccc agctactcgg gaggctgagg 144060
cgggagaatc cgttgaacct aggaagcaga ggttgagtg agctgagatc gtgccattgc 144120
actccagcca ggaacaaga gtgaaactcc atctcaaaaa aaaaaaagaa 144180
aaaagaaaaa gaaaagtggc atgctcatag caagtactat gttagtata gtaaatacat 144240
ggatcttggt ttcataatac aaattaatta atgatataaa gcattcatta aatatctcaa 144300
tatttggaa cttttctttt aataataaag tctagtaaat gttgttattg aatagaaata 144360
cacttttaaa aacgttggtc attattttta aaatatatga atcagaaaag gaaaggtgat 144420
ggaaaactca tacaaatata gtaaatgca ctgtttttgt gaagatgatt ttatgcagaa 144480
ttcaaagtag atgcaattag tgcttcaaat tagaaactgc aaatgatgaa ttcacaagca 144540
catgatttaa caaacaatc agagaagaga aaaagtttct gatgaatttg agtttctaca 144600

```

tgtggatttc	atcttgtgat	aagccattct	caggagaaag	catgagatgt	tattcgtcat	144660
tccagaaagt	gggagaacat	aaaagttgag	aacagttatt	gcataagcaa	gaattatcaa	144720
tgaacacaat	taatttaata	catacacagg	catttgactg	tatgagaaaa	agcatttcat	144780
gttattttgct	ttgtagatac	aatatcagtt	tcgatgagat	tcattgctgca	cggatgggta	144840
cacctgtcac	ctgtgggtgg	gtaattctta	tacctgtgat	gacaagtaga	ccacggggga	144900
ggactcgcaa	aggattttgag	tgaggagcca	ctgatgctgt	gtaaatgcca	gagccgtgac	144960
cgtgcacttt	tggggggatg	tgcttaata	tcttcagatt	tcacagaagt	tagggcatca	145020
acggaaaata	cttttgtttt	ttctagctca	tctgcacaag	ttgaacacac	attgttagat	145080
agtgtgtatg	tcagagcata	tgtgaacaaa	tgaaggcgaa	aagtcaatag	ccatacattt	145140
gtatttgtat	tgctctttat	caatctaaaa	aggtaacatta	aacaataaaa	tgacaattca	145200
atttttaaca	gcattctaga	ctctagaagg	tgttttctaa	tccactctct	taagccattc	145260
tcacaacagt	acaagggaga	atttcaatga	gaccaggcca	cccagcaata	tggtgacaca	145320
gatgagacac	ctcctcagat	ttttgtttca	aaatttttgt	tctgcctact	atgccataga	145380
tgctgttaac	atttaattag	aatttatatt	tttaataaat	aaaactaaag	agtttttgcg	145440
attgtgttaa	cactcacgtg	ttaaatttgt	aattatgtcg	gtaaataaaa	agagttttaa	145500
tccacacccc	attagccagg	cggtggtgtg	cacatctgtc	tgtaatcgaa	gttttagaaa	145560
actgaagtgg	gaggatcact	gagcccagaa	gaggtggaaa	ctgaagttag	ctatgattgc	145620
atcactgcac	tctagcctgg	gagacagagc	acgacccctt	gtccaaaaaa	agaggaaaaa	145680
aaaaaaaaag	aaaacaaaaa	ggcgcaacac	agagagaagt	gatcttatcc	ttatccttat	145740
gttagatata	tatttttttt	ctggtcagag	aattacaagt	ctgtctagggt	aggctccat	145800
ttgagaagag	aaaatcctgt	tctattgact	tagagagtta	aagaccctgc	cataagcaga	145860
gagggcacac	gtgggactct	gcgccaggac	aggcacaact	gtgggggacg	cttcatagca	145920
gcttttgcag	atctcctctg	cacctgaagc	gccgcctgta	gaagctagat	gaaacctcac	145980
aatgccaaag	ggttgctggg	cttaggaagc	acgggctgca	tagggataac	tagatcttta	146040
tggtacctga	agccactgcc	ctttgtctgg	atatctgtct	ctttgtgtgg	ccttataggc	146100
tgcatttttc	ttgcctgtat	gcctgaggtc	ttgatttagt	cacactggca	gggtacaaac	146160
tgccctgact	ctgccagtga	cttcacagtg	cagttgcaat	gtcatttagt	taccttccct	146220
tgagggaact	gctactgttg	atgaagagtg	ctttggacca	aatattttat	gggcccacag	146280
ttcatgtatt	aaagccctaa	ttcctaattg	gatgatgtga	ggaagtggga	cctctggggg	146340
gtaaaataag	tcattgaggg	agagccctta	tgaattgagc	tagtgtccat	gtaagaagag	146400
atggaagata	tgattttctt	ctccaccatt	tgaggacaca	gcaagaaggc	agccttctgc	146460
aagtacaggaa	gagagccttc	accaagaacc	ctctctgtca	gcactttgaa	ctcagactcc	146520
ccagtctcca	gagttatgaa	aaataaatcc	ctgcatttta	ggccaccacg	totaagacag	146580
agaggtaaaa	tatatagtgc	ttgggccact	cgtttgttat	taaattcact	atgtattcag	146640
cttcacttat	atccttagaa	aagtttgtga	atgcagtga	atgggtggag	aacggtagtt	146700
gcagataatc	agagaagaag	agtagagaca	gggtgcaggg	gaaagttagt	cctttaaatt	146760
gaaaaatgat	aacatgatcc	attcatataa	atgcactttt	gttcaattct	gctccagctg	146820
tattatccag	cttctctaaa	gaatgggtgc	tgtacaggac	aggggtgatc	ttaccctgat	146880
gaaagatggc	tctccaggca	gatagcagga	gagacattcc	taattgattc	caggagggaa	146940
tgccaccacc	atcaccacta	tcatttatgt	tgatatacaa	actgaaatgt	caatgatcat	147000
ttctgttctc	taacatttcc	aatcagataa	tttaaacatt	ttactctaag	agtttttcaa	147060
ctcaaatggt	atgcttctcg	ttttgtgcat	ttgtattatt	aactctagtt	tcattgggag	147120
cccagttttt	atacatcact	gtatagcttt	gatgttaaca	tgatttttag	tttaacataa	147180
ttttgccaaa	gtgagtatta	acctagtgtg	atcacacaat	ttgggttatt	tttattctga	147240
tgctaaggga	atccaatgat	ttttgttgat	agtaaccctt	caattctatt	gactgtaaag	147300
agtaattctag	aagaaattta	aagaaaattt	aatcagattt	tacgttttct	gtcgctaatt	147360
tggcatctta	atagctgttt	attttattcg	agggtgcatc	tgagtcttgt	aatgttcaca	147420
ttgcaatatt	aaagcataac	agatgagcct	aaaatgttca	gtgggttaatt	tgtacagatc	147480
aagatgttgt	aattgcaata	accttaattg	ggctaaccct	tgaataacat	aggtggaaac	147540
tatgtggata	aatccactta	tctgcagatt	tgtgcccacc	tcggttacc	cagacacaga	147600
aaaaccaaac	cctcctcttc	ttcccccttc	tcagcctact	cagtatgaag	acgatgagga	147660
tgaaaacctt	tatgatgac	cacttccact	tcagtgaatg	taaatatatt	ttctcttccc	147720
tatggatttc	ttaatagcat	tttcttttct	gcagctttat	tgtaagaata	cagaatataa	147780
atcatacaac	atacaaaata	tgtgttaact	tactgcttgt	tatctgtgag	gttttcatte	147840
aacagtaggt	tgctcagtag	caagcttttg	aggagtctaa	gttacataaa	aattgctgac	147900
tgcgtggctg	gggtgggggg	agttggcagc	cataacctcc	acattgttca	atgggtcaact	147960
gtataatttt	ttagcgggaag	tgtaggaatg	tttagtatga	tacaacctag	gagaaatatt	148020
ttgtataaat	atgtgtaaaa	tttaagactgt	agaactaaaa	ataatacact	tagttcatta	148080
gttattaata	gctaattatt	atttagctct	gtgataataa	attgacatgc	attacctcat	148140
ttgtcacatt	gtcatgagat	caggattttt	atttgtcttt	catttacaaa	tgaggaaact	148200
aagtctaaga	ggagtttaagt	gatttaccca	gtattacatg	gcaaaaaaaa	agtcagagtt	148260

taaacttaaa	gccgagtcctt	tgcaattcta	aagtcacatgt	tcttagctac	tgcctttctg	148320
aacttattag	tcaacctgaa	ttctgaggat	ctcacaaaaa	gttacgtgct	aattattttg	148380
gaagcattta	tattatgagt	ttctaaaatt	acaaaattctt	ttatgtatga	tagggaggta	148440
aaatcatggc	atttagaggc	aaacatattc	tatttagttc	ccattttgtc	tacttaacag	148500
ttgtgaacct	gacgagtacc	atttagtaac	ctcaactgtt	aagggtgtga	aacaagttca	148560
tttcatagag	gtactgggga	aagttaagcg	agagaatata	tgactaacia	aaaatatgat	148620
tttcattttt	tccttgtgtg	tatacttaag	gggaaattta	ggactgccta	cttggttatct	148680
tccactgaaa	acaaaaacaa	acagaagatt	aaaaatacaa	acatgttgcc	cttggtttatt	148740
ttgtctaact	caagtgttgc	tgatatcatc	catttctgtg	taaccacaag	cacagttggg	148800
aaagattgag	ttagcaatgt	tcctcaatt	ccagattgcc	ttctctacca	tctaatttct	148860
ttcagtggta	ccacaattct	tccatgctgt	aatagaaact	ttgaaattat	attttgctat	148920
ttttctcttt	tgcatcatta	ttttacattt	tggtgtcaaa	tattttattt	ctgcagaata	148980
tcttataatc	ccatgcttct	aatttctagt	cagtgaactt	cttgattctt	attcctctaa	149040
tgtacaatcc	aaataacact	ttcttctctt	ctttcattta	acaagtaata	atttaacgtc	149100
aacactgtgt	catatctgtg	gtgtcaccta	tcagtatgca	gaacaccaag	ggattccac	149160
tgtagtgtga	ttcttcttct	tggtatttaa	caccaatgct	ttcttggcct	cttaacgctt	149220
cctccttgat	atttttcagt	tccaggcact	ctgccacaca	ctctcatggc	cacacacagg	149280
accttggttaa	cacctggaaa	ttctccatct	ttgaactatg	aaagattagt	atctgcaatc	149340
agtaaatggc	atcatctctc	atagctttta	ttccccaatg	ccttctatto	agccttgaac	149400
tttgtttttt	ttctgctgtc	tgcttcttcc	ctcattaagc	cagccacatc	attcgatccc	149460
ttcactcaca	gcattttcaa	gaacatcccc	tcttctctct	tttatcttca	gtcactctgg	149520
tgtgaatata	ttacaatta	tacaatacac	tttcagtgtt	cctagactcc	tggtgttaga	149580
gacagctgtg	taacagaatg	ccctgggtgca	gactgggggt	ccctggatag	cgagggtctc	149640
cagtttcagc	tggtctctct	ccttaggtaa	gagatagaaa	caagaatctt	agtattattc	149700
taatgttcaa	gtatctgggc	ttttctgcta	tagttttccc	ccataaatct	tgatttttag	149760
ccatattaag	tggttttttc	ctttaccaga	taaacaaatt	tgctctccaga	ctttctgcca	149820
ttttcacttt	ttttacotgc	caaaactttt	gtgtctttac	gtactcaatt	taaaatcctc	149880
ctttatgaga	tttcagcttt	ttctcttgga	ctgagattga	gaacctctct	ctgggctcca	149940
ataatttcaa	atgtctgctt	aaattttgaa	gcataatttc	attttatgag	aatacataga	150000
tttcagacta	ctaagcaata	gaaatatgga	tacaagacat	acacggggacc	acttggggca	150060
tagtaagtat	tgaagaatac	tgaatacttg	ggtagtgata	ttaggccgtt	ttaggccgtt	150120
ttcacaccgc	tataaagaac	ttcccgagac	tgcgtaattt	ataaagaaaa	gaggtttagt	150180
tgactcacag	ttctgcatgg	ctggggaggc	ctcaggaaac	ttagaataat	ggtgaaaggg	150240
gaagcaggca	tgtcttacat	ggcagcagg	gagacagaga	gagagagaaa	gaaaaaaaac	150300
agccacttat	aaagctatca	gatctcgtga	taactcactc	actgtcacaa	gagaagcaag	150360
gtggaaacca	ctcccatgat	ccaaacacct	cccactgtgc	cccttcacca	acaggtgaag	150420
attacaattt	gtgattacaat	ttgagatgag	atttgggtgg	ggacacaacc	aaaccacatc	150480
agggagtatt	ataattggta	aagaatgact	gtccctagt	cattgaaatg	taagtctcat	150540
ggaactaacc	tatatgtgtt	taatgcaaat	ttgaattcaa	gatggcagac	taaacatatg	150600
cattttttac	ttatctttcc	aagtctttgt	taagttttta	acggattata	taaaagagca	150660
agaagacaag	aattaaaaaa	gaagtagtaa	gccacatggg	aagtatgcat	gagagttaga	150720
agctgaggaa	gtgagtagta	atctaaaatc	tcagagaggt	tttgggttca	taggtgtgca	150780
aaaagagaat	gagcattggg	tccaaaaaac	aaggcgaata	atggaagggt	aatagttgaa	150840
cagttgcgct	gatcagtcct	gctttgcagc	tcattgtcaa	agaagagatc	aagttggaat	150900
ttttggccaa	atactatagg	gttgtctgtg	atatttaata	gtaactgagt	gagcagcctg	150960
agagtcaaga	ctgttagcta	tgttttgaga	ccccaaagta	aaccccttgc	attttatcag	151020
ttgacagaag	agaagtgtc	tacatgtgcc	ctgacttaag	cttatttttc	attcgtaagt	151080
gttaggtaat	ttttagtaat	atcattagcc	ttctatatgt	agagttatgt	gagaattaaa	151140
ggtctctttc	cctgcaatct	cagacagaat	tctactgtac	atgcaaacat	gtattatcaa	151200
gtatggacaa	aatcaattgt	actgagtata	atgaagtcca	ggttctctgc	cttttatgtc	151260
ctctaattgt	acacagcata	aaatgttgac	tcaagccaga	gcacaaaata	ctcaaaggac	151320
agaattcagc	ttatcttatg	gtcagatatt	tgactaaaaa	taattgaatc	agtgattgga	151380
tacaaagact	tgagtagcat	cacaaaagtgt	tcaggctctt	gaacttccta	ggtccttata	151440
accttctttt	gaaacaatag	aggccgggca	tggtggctca	tgctgtaat	cccagcactt	151500
tgaggaggcca	aggcgggtgg	atcacgaggt	ctagagatag	agactatcct	ggccaacatg	151560
gtgaaatctt	gtctcttcta	aaaatacaaa	aattagctgg	gcattggtgt	gtgtgcctgt	151620
agtccagct	actcggaggg	gtgaggtagg	agagtgcgtt	gaacctggga	ggcggggggt	151680
gcagtgagtg	gagattgtgc	cactgcactc	caacctgggt	acagagcgag	actccaaaaa	151740
aaaaaaagaa	aaaaggaaaa	ataggagtac	atatggaatt	ctaggcatct	cacaaatctt	151800
taacgttagt	atttcttcaa	ttaaaatact	ttttaaaaat	tctaaccgct	agaccactag	151860
agaaaaatac	ttttggtttt	attaaatac	ttatttaacc	aacctctggt	ctcatcagag	151920

agaaaaatag atggcatggt tttacotttc cttgtttcag gcttctgatt agaattccta 151980  
 ctagtcaacc atttccaggg tgcacattct taatatttct caaaatttat attaggtcta 152040  
 aacctacaaa atttattttt ttaagatcaa agaattgtcaa atattagcaa gtcacatggt 152100  
 ttatactggt aaagcaaaact aaatatggcc tgagaaagac tttgtacttc tgtatttgag 152160  
 accttgtgga tgaaccgcaa cctaaattaa taggcagaga aaatttttaa ctttaacttag 152220  
 gaggtagcac ttataacaat agctgggtct tggtaaatcc cagcagccat acttcagcca 152280  
 ctcatacacc gctgagtatt caaactgtgt tcaaataagg caaatgccaa tctgtaacca 152340  
 gtccagctgt ttctgtacct cacttctgat tccgtatgt cacgtcactt tatttgctta 152400  
 taaatttggt ctgaccacaa ggcatccctg gagtctctct gaatctgctg tgattctggg 152460  
 ggctgccccaa ttcatgaatc gttcattgct ccattaaact cctttaaatt agtttggctg 152520  
 aaattttcct ttaacaata ccatctctcc tgggagagtc aaactcatat ggcaggtgag 152580  
 gattcagaga gattgttttc caaaagatag aaagttaaagt tgccattttc ttaaggtcta 152640  
 ggcagagaaa ctggtatagc tgtacttcta ccatattcta tcagttacac actgtcatgg 152700  
 agcaggctct tatgcaatgg aagaggacaa aggtggcctc atcttccctc catgctctgt 152760  
 tacaatatat aatctgtctc cctttgcatg ccctgtcttt ccttggattt taggctattt 152820  
 catttccctg caatatcagc tctctgttgg attcaagaat agagatgatt ttgtagtatg 152880  
 ttccactatt ttcccccaa attattaagg tattaagtaat gctctttcat gttccatata 152940  
 ttaagtggaa gtggaacttt cttgctataa cttttgaatt agcaattcta gactctagac 153000  
 tccaaaggag tctgtatctg caaaagtctc tagggaaaat aattgtgaag tttaaattct 153060  
 aaaaaagaaa aaatccttgt ttaactaaat gatcttcta aaataacaaa aaatgatatt 153120  
 ttataattat aagaactaag gaaattgtct ttatataaat gtctggagaa atagatttct 153180  
 gagaatatat ttcagcaaaa caaggaaaaa ataaccagg gatatactgg gactgaagta 153240  
 acagtgcctc ttaccccgaga gataagtgc atgaggtccc agaatactt taagcaggct 153300  
 tagaagccga atgtataaaa tatatcatgt aaaaataaac gtgacacaga gggcacttag 153360  
 ggtcattctg attcttgacc ttgatctgag ttccaatgac gaaaatagat tggccaaaag 153420  
 atgacatact agccacgcct ttattttcat tacatacact ttttactcca tactgcacag 153480  
 atatatacaa ttattttatc ttccatacca ggcaataagt tacttgccca tcttcttatt 153540  
 acaataatg tagataaagt atcttatgaa agtagagaaa tgtagtacgt tatttgataa 153600  
 gtgtgtgagt ttgttcatat ggtcttatag gaaataacaa acataattta aattggtaca 153660  
 ggtgtgtacac agaaccattta aaaaagtctc acctaatctt aaaaccatgg atttttactt 153720  
 aatgtttctg cattttataa tcattaaaaa caatggaaat tttgttgagc ttaatatgaa 153780  
 ttaaaatttt tatattagga tatcgtaggt tttgccaata gaattctaaa acttaaatat 153840  
 atctaatttg attactatgt aggttatgca ctgtgactca ttagaataata tgatgtgttg 153900  
 ttaaaagtca taatgatcat cagatgaaat atggccatct ttaaagttat ctcatagatc 153960  
 taagtagtga ataaacgaga caagagtaat gtaaaaatgg ttgacattat caatgaaggg 154020  
 tcaatggata actgcgtggc tccggttctg atactcagaa ggaaaccaca tcacttaagt 154080  
 agtaatatga ggaaaatgaa ccacattttt aggaaacatc acacaaaactg taaaccaaga 154140  
 acactttact attaaacttt ttttcaaata tgccaatgcc atgaaaaaaa acaaagagct 154200  
 gaggacccat tccagtttaa aggagactca agagacatta cgaccgaatc ctctatgtga 154260  
 tgtaacataa gcatcaaaag acattaaact ggtcaattca taaaactgaa atctggaagg 154320  
 tagatcaaat aaaggttttc tatcaacatt ctgtttcctg aaacggtagc ttgtgtttaa 154380  
 caaagaggat gcccttgtag ttagtactaa ggggtgtagg gtccgtgtgt cagtgcactg 154440  
 cacttggcag tgagtgcgtg ggcgtacgag catgcacacc acacagtgca cacaggggtg 154500  
 atgcaacagt cagtacatcc aggaaaagga tttatggcag ctccgagtga gtgcgtgggt 154560  
 gtacaagcat gcacaccaca cagtgcacac gagggtgacg taacagtcag tacaccagg 154620  
 aaaaggattt aggggagctc tgagtgtgtg gatgtacag cttgcacaca cacagcacac 154680  
 acggagatga tgcaacagtc agtacatccg ggaaaagcac ttatgggagc tctgtgtact 154740  
 attgttgcaa tttttctgta agcttgaaat tgtcctcaa taaaagtga aaaatgttac 154800  
 cccaggaagg atttggcaca cattatttta gatgaatact togaaaactt cctttatttt 154860  
 ttgaaagtta ctggaacagc ctgattacac atttgggatt taaccctct gccttaagta 154920  
 taaacttaga cgtaagtgtg agaaccagtt ttaagttatc ttgagaagca caaaatttga 154980  
 gttactacac agtttccctt taaatggaaa tgaaatacat ttcaagtcact actaatatat 155040  
 gttgggtttg aactacttgt agcccttctc tagaaaacta aaacaaaaac tctcttttaa 155100  
 tttcatgact ccttcaaata cggaaaactat gaagatttat gtcttatgaa gtttatttta 155160  
 cacataagtt attagattct tttaaaatta ttattattat tatttatttt tagagacaaa 155220  
 gtctcactct gtaaccaaga ctggagtgcgt gttggcaaat catatctcac tgcagcttca 155280  
 aattcctggg ctcaagccac cttcttgctc cagctcctct agttgctggg attacaggca 155340  
 tgagctacca tgctgagcat acaaatattt tgggtctgat aaataattat tgataaaatc 155400  
 aacagaagtg atcattataa gttaaatatt tacagtgtga tttagagaac atacctgtag 155460  
 aaaggattga aacaataaaa tattatttggc ttggaatttg tatagaattt tgcccttcaa 155520  
 tatagactgt catctaaaca ctcaaataga agacattcaa atttacaaaa tgggtgcttta 155580

ttggccagcc	tagataattt	ttcttgcttc	ctcagccccc	tttttactta	ctcctgaggg	155640
cttaatatca	cagacaactg	aaattcatct	gcactctact	tactatgtgt	aaccagctgt	155700
gatttttatt	tctgatgcta	ccacatgaaa	aaaaaatttt	tttcctgagc	tatagatgtt	155760
aattgggtgc	tttgatctct	gctaattaag	ttctagcaac	accttctccc	tcttcctcgt	155820
ttgcacttaa	tgagctcact	tggtgtatta	atttataaga	taattagata	aaacaaaaaa	155880
agttgaattt	cttcaggagt	ggctttggaa	agtgcacctt	aaatattcaa	acatggataa	155940
aaatgtgtac	attacaagac	aaaaaaatca	ctttaagaag	aaatatactt	aatgttagcc	156000
tgtagactca	tactgacatt	ttaaataatg	tgatgatcag	ggcttataca	aaaaggaaga	156060
gattgattca	aaagctcagg	caatcacaa	tattttgggtg	taagcaaaac	aggatcacat	156120
atggtacaac	atgatgctca	taccagaaac	ctttgcaata	tttaagcaca	acaaacaaat	156180
tagttaagca	atgtgcccc	tggaactctaa	ttaccatttt	tatttgcatg	ttcatttttt	156240
tctaaaattt	ttttaagcca	gtggccaggc	agtattaaat	tctttgctgt	ttatcatgat	156300
accagaattt	atttgacaaa	aagtggaaaa	atcattgacg	tacagtgaaa	ctgatcattg	156360
tttaatagta	gaattctgtc	gttcatcagc	agttagggaa	ttgtggaaat	tgtttttagt	156420
gagaacatcc	agattgttatt	gattcccaat	ttctgggtcca	actcaaaactc	atttgggcca	156480
ttagctttcc	aaactgtcca	atatcatcag	aaaaggcaat	ggcatctctt	actgctctat	156540
ttttgtcttt	ggttgagtaa	tggaattccac	ttgactggca	tatatattct	ggctgcccct	156600
taggaaatct	aagcatttga	tgtcaaaaaat	acacatcttt	taaaataactt	cttaatatgt	156660
aggtaaaagca	aacattcaaa	attcaaaaata	tcaagctgta	aatgttaata	ctaaacagga	156720
acatatccaa	tcgacattga	aattgtatat	aatgaacaag	aaacaataaa	ttaaagtga	156780
aagtcataat	aatatgcaat	atttcaggta	ttctttgata	gatgtaatga	cagaagctac	156840
taccttgctt	gccccagtg	tagtaacatt	gagctttcca	gctaggatta	gagcattttt	156900
atatgaatcc	ttaaaaaatg	caataactac	tcctccagga	gaggatgctg	tcctgcttct	156960
ctgagggata	gcactaatat	ttccagggtg	tctgctttac	atcaagactt	tgcttttaga	157020
tggaaacagg	attgtaggga	gactgtgagc	tcaacagtaa	gattagtttg	tctccatttg	157080
atcttttgat	gaatgcaatt	ggggtaaaaac	ctcttcccag	tgctattagt	agctcatgaa	157140
ttagaaaaga	taatttttct	cctacatatt	caagaagaat	cagaaccagg	taattgacat	157200
agccaactaa	acctctatag	ttattgggtt	tgttctgact	caaatgctaa	cttgaatccc	157260
ttatgactgt	taagagcata	aagggggaaga	gaatgcccag	ggaggcaact	ggagagacag	157320
agacagcaga	gctgtgagaa	gactgtttgc	agaaaatatt	cttcccatgt	gtggacttgg	157380
cctttgaaag	agagaattac	aacaggaaat	taggaatagg	ctgggtgtagc	ctcataaaca	157440
agtaaaaaat	tgtgtcagca	cacaaaaata	tttactaga	ctgcccttcc	cctaggaggt	157500
tcacagggtg	tataataaat	gaaatcagca	atccttctac	atgtctcgct	tatatgggta	157560
ttttggctca	gcogaccatt	agttatttgt	aagctggggc	ataaatacag	acactagttg	157620
aatattttat	tacagctcca	gaatccacct	tcactaattt	gtctacttat	ggggaatata	157680
gctgtggggg	gcatggaaaa	tatttactct	gggagcccct	aaatgtaatg	catgaatatg	157740
aatgtgctct	tggaacggga	ggtcctggaa	ctcttggtag	gtgttggtgc	cctactgacg	157800
ctttatcatt	ttcttttttc	cttcaactga	atataaaaact	gaaaagccgc	ttttaacagt	157860
tagtaagtgg	tgatactgcc	agactcatcc	tataagaagt	tagctaaatt	ttatccagac	157920
ataatatttt	ttattagctt	acttgttagag	tcctcctcca	gagcccaact	aaacgtgggtg	157980
ctaaacattc	ttttgocctt	gtagaggaat	aacattctga	gactcctttg	ttttttcacc	158040
aaacagtaaa	attaaccagg	tattgaattc	atactctcat	taaagtcatt	gttaccattc	158100
ttttacaaaa	aaaaataactc	aaggagtaaa	gaaacataaa	aattgacact	aaaacaatat	158160
cacctacgtt	catctttctc	tggaatgttag	atttggaatg	tatgcataga	aaatggctct	158220
agtgtcaaa	atttaaatgct	gcaagagaaa	tcactctcaa	aagaatatct	cctttaattc	158280
tgtcttctta	tttatcaaat	aatatctagc	gaccgcactt	ttcttttgtt	ccttttgtat	158340
aagaaggaat	tatagcgggt	tgccacatta	gagtaactta	aaagtgcctc	ctaaaatttg	158400
gtagagaatc	ttcttaaaat	tttactccac	cctgcactct	tctggatata	tatcactcct	158460
gagatgggtt	ttccaggctt	agattagttt	tctaaatctc	tcagctttct	tttatctcaa	158520
gtgacacagc	agttaaatta	catcatgagt	ttagcctgga	ctctcttgag	cccccaaga	158580
aggaaaactg	aaattctgca	aatgaacct	tatttaagg	tgaagggaat	aagagacatc	158640
tcaagccgtc	ccttatttag	aaggtaaact	aggaatcata	aaataatata	tttgaccgag	158700
tgctcagaaa	tgattaaatc	caacaggctt	gccaacagag	ttgtaccgta	gaattgccaa	158760
gggagttatc	ataaagcaca	gatactgggc	tctgcctttg	accttctggg	tagataaatg	158820
ccagtgattt	tgaaagtaat	tcctgtgat	tctactgaat	ctccacaaat	cttggtgaac	158880
ctctcaattt	taaaaagaag	tttttttttt	ttttcagttt	aagccattta	ttcaaagtca	158940
aacatgcttt	agtgatggag	aagaatctgc	ttttaaatgt	attaaaacaa	attgatagga	159000
atcattgttc	tattgttttt	cctgtataaa	gttagacaaa	taggatattt	ttgaagttat	159060
aaatcattta	aataccctta	aactgtttac	aatatcggtt	gttatatttt	tatcttctgg	159120
ggtaatgtta	tttaaatata	acgaggtact	gagtgcctgc	tgaaactaac	tagaaatctt	159180
aggccttaaa	aatagctatc	tcataatttt	aaaagaaaca	tcaataattt	ttatatattt	159240

ttaaatttat	gagatggaaa	gttagctttt	actaatctta	tcaggctaaa	ctgagtaata	159300
atgaagaaaa	tgaagagttt	agaaatattt	tttggaatat	cagtgaatc	tctaaaaaga	159360
aaaaaaaaac	cctcatctat	atgtgtgagc	gacataaagt	cacactttgt	tgaaatgcaa	159420
aatggccttg	gaaaatttca	tggaatgaaa	ctaaatggca	gcacatcact	tttccagatg	159480
gggtcttctg	gttgtggctc	agatcatggg	ggatgatggc	aactcttoca	aaggaaaaat	159540
agtagcaatc	cagtcacaag	atagtaccaa	acatagactc	acagatgagt	aatatgtcag	159600
tttcttaaaa	gaaagagttg	ttaggaatga	gagtaaggat	gtgcctcgat	tacaataccc	159660
cagatacgat	actattctga	aaatgggtgt	gggtttttgt	gagggatgcc	aaccacacat	159720
tgtgctagaa	atctgcatcg	ggaagaaata	ttttttgcta	catgttacia	atatgcctat	159780
agactgtgat	tcaaataaat	aatgtgttta	ggccaggcgt	ggtggctcac	gcctgtaatc	159840
ccaacatttt	gggaggctaa	gccagggtga	taacctgagg	tcaggagtgc	gagaccagct	159900
tggccaacat	gatgaaatgc	catctctact	aaaaatacaa	aattagctgg	gtgtggcgac	159960
gggcacctgc	aatctcagat	actggggaag	cttgaagcag	gagaatcgct	ggaacctggc	160020
aagtggagat	tgcagtgatc	tgagatagca	ccactacact	ccagcctggg	tgacaggatg	160080
agactccatc	tcaaaaaata	ataaataaat	aaaaataatg	tttttataat	tttttcatag	160140
aaaaataaac	ctgaagttag	gcagtgcaga	cttgaatttc	tcattctcaa	ttcccattca	160200
ttttcctgtg	tttttgaggg	tttctcctgg	aaagtataac	aattcaatgc	tttatcttct	160260
gcccttcaga	agtcgggaca	tcacacacac	atttttacctt	taagatattg	gactggattc	160320
atagaggaat	gagccatata	caacatacat	aggagtatga	agaaaggagg	acagccttca	160380
aagctaattt	tgactactaa	attgtagttt	aatgtaatga	gaaataaaac	tagaagatga	160440
ttgatagaat	gcagagggca	ctttaaatca	gagagacgat	ttgaaagaaa	gaaatatttt	160500
tatataccct	gatttattag	agatagatta	aaaagtctgt	ttaaaaaacg	tgatcaagaa	160560
cttgacgtga	taagttattc	tttttcaaat	atcatgacga	taatgctgca	taatcaccca	160620
ttcaaaaatt	tcacggcttm	gaacaacaaa	tattactctg	gagctctctg	ttctgtgggt	160680
ggctgagttt	tggctactct	tgacttggct	gagctgggag	gatggatctg	cttcaggctg	160740
caagctggct	gagcttggct	catcacgttt	cattacgagg	cacagactga	agtgtttgga	160800
gaagcccagg	cagtggcatg	ttctttaaag	gcactgagag	gagcttatga	atgaatgaag	160860
cttaactaca	taagcwact	tgatgaagtc	aaaatgcaag	gactgggaag	tacatgccac	160920
caaccatcaa	gccatggtaa	agctgtcgat	gtactacggg	ggagtggagt	gctggtagca	160980
atcattcagat	ctttcacaaa	acactggcaa	aatgcagggt	attcagacag	aatcgtttaa	161040
tacaaaaaac	tacttaaaac	ttaagggtgt	attagctatc	tgtctccaaa	tgtacttttt	161100
gaatttaagt	tataaggact	ttcactgcag	aaagtatgca	ataatgaaga	agataattgc	161160
tttggtcatt	gggaaagtgt	ttttcaggta	gaagtgaatg	cagaaacaac	cattccaagc	161220
aaaaccaatg	acttccaggg	taggctggaa	agagaattct	cttaatagaa	aagaaaagag	161280
acttggaaat	gagatagaga	cgctaaactg	agctactcct	gttggggctt	tgctgatcct	161340
tcattggaata	cactcattaa	ttatcacaga	agccacctat	gtactctgca	ttcattggga	161400
tttgatgtat	aattgggtct	taacttctct	atgttttagca	ttttgcttag	aagttagcac	161460
ttatcttcag	gtattataag	gcttttaaaa	tgacattttc	ccctctagag	gaagtatgcc	161520
tttgtttggt	tatttggttg	ttttgtcttg	tttcttagag	acagggtctt	gctctgttgc	161580
ccaggctgga	tggaagtggc	ttcatcatag	ctcactgtag	ccttgaactc	ctggactcaa	161640
gccatccctc	tgccctcagtc	tccaagtagc	ctggggctac	atgcttggct	aattttttat	161700
tttattttat	tttttttagac	atggggtctg	gctatgttgc	caaggctgat	ctcaaactcc	161760
tggttcaaaa	ctcctggact	caagtaatcc	tcccacctga	aggtgggagg	ccagcacctt	161820
gacctttcaa	agtgtcgaga	ttacaagcat	gagccgctgg	ttacaagaca	tggccagaag	161880
tatgggggtt	tctatgcttt	cacatataat	acatacttag	gttttctcga	tgatgaaatt	161940
atgtttgggt	tacttttagt	aaacogtatg	tgctttgaat	taataaaaaa	gttgtgattc	162000
atttattttc	aactgctact	ttccttctct	tttggagggt	ctccgtttaa	agattccctt	162060
tatgttgtgt	cacagagtct	gaattgacac	aaaagcattt	ccaggcatta	ttcttgttcc	162120
cggttggctg	acatagagcc	ccactcactg	atggcactgg	ggaaggacag	agcctgttgg	162180
acaaaccaac	cctaggcatt	agtatcaaat	ggacaaaagg	atcctaagac	gtttagacaa	162240
ctgttgactt	ttatgaaact	gagacaaaaa	tggttaagaa	tttagataaa	tatttccctt	162300
tgtaaaaaat	acatgaattg	taagtttcct	acactgtaaa	ttagaaaaat	taaatatat	162360
aaaactatat	aaaattataa	tttttttctt	aatattggca	tttttatgtg	tactaaggta	162420
tttacctcac	ttttagaatc	tcgaattcta	ttgggaggtg	attattttca	actcataaat	162480
caccttttaa	acatttgatt	cttataactt	ttccttctca	ggaaaaactga	acctatattt	162540
atagaatatg	gatgctccag	cagctgggtc	aaaaaatgat	tccgtagtaa	taatcacatc	162600
aattatatgg	ttaaaaatgtg	aagcctgagt	caccgaaaag	ttaggtcatt	taatttcttc	162660
aatgcagttt	ctaaagtttt	caagaagtgt	attagattta	aaagagactt	gaaagggaat	162720
ctgatacaat	gtcattacaa	ctgagaattt	tctcaccaga	attcttgaaa	gtactgaagg	162780
atgcaaccct	agtacatttc	tcattacaca	tgtagatcag	aaaccaagaa	gtctatgagt	162840
aacgttttga	aagaacaatg	tttaaatatg	aaaaatgatg	tgactcatga	ttgaaaaaga	162900

aagtcacgtgta aattggaagt gcaaaaactt tacaggagaa atctatgaga ccaaaaatag 162960  
gtagtaaaat aaaaatgaaa gtaaatatgg ttgcctaaaa cacattaaca aatgaaacaa 163020  
atagataaaa ccaatgagaa tgggcagcac taagagtaac agagaagttt gacccaggca 163080  
atattgatga attacttggg aaatctggta ataatcaggt aatatgggag cagcatgatt 163140  
agttaataat tccactttaa aaaattcaca cagaaaagga cagagtatat atggaagaca 163200  
tgacacagctt ctatattaat gtgacactgc tctgagtaat taaacaccgt ttccaggaaa 163260  
gatttggtca gttgttccta gaggggtgtg atacatccag gacattcagc ttcatgactt 163320  
ctcatttggt ttagttttta tcttttaaat tgggggaaaa tgatgtgtta ttcactgac 163380  
ttttcttctg agatgccaat ctaaatctgt tatcttatta aacaataagc agattttgtt 163440  
ctgatttagg caattaccat ggtagtaaa caagaaaatt accctaaca tgggaattagg 163500  
ttttgtggac atgatcaggt catctttttg aggtatctgc cttttcctgc attcataata 163560  
accactatta ctctgattac ttttatactg tccccataag tagtgaggtt ctcaactacg 163620  
caatgaaaaa cagaattcag aggaaatgta accttctgaa ttcttattgt gcaaatttcc 163680  
tcttcattca gtgccatgct caccctggaag aataaatttg tcaaatattc ttgccaacca 163740  
aacagacgtc atattccttc agagaatttt cacactaaaa acaactatta gttttacgct 163800  
gacagacaaa gtcaatccta atgtctttgg atcatgaatt tatctacaga ttccagaaga 163860  
aacaagacaa acatttttgc caatgtcatc catcaatcaa gtggcaaagc agatgtgtgg 163920  
atttcaatgc agatagtttc tgtccagctg tctgctctcc atcttgatg ttggttttgt 163980  
tgtagggtg gctctcctca taaacttgag atgtgtcgta gcaccttgta aggtatgca 164040  
gttcttctct ggcatccact gaaaaaagag tggatatctt ttctttagga aattcatagt 164100  
tgggtattca gtgcagctg cattcttaaa caatcactgt tgccaataga aaacaatata 164160  
ttaactggct caccataggat cctatacttg tagacagttt cagccaaata caaagatctt 164220  
aagaaagggg ggattggtt ctctacctca agaaaaatta aatcaatgtt gatgcaagag 164280  
tgggaatgca tgttgagagg caaacagtaa acacaacaaa atgataaatc agatctgctg 164340  
gccctcaata ttttctttt tttgccaatc cagaaaaatct tctatgattc agagctcact 164400  
cattaaatgc attgcccttg gataaaagcc attggtaaaa taaccaacaa gtcactctcc 164460  
tttgaacctt tagccatatt catgtctacc tacattatgt atatttctt gcttcccaca 164520  
ttccatcaga gtaaaatact tgcgaagagc cagccatgtt ctaggacata tggacatgaa 164580  
cagatttata ggatgtggtt ttctcatccc aactcttgaa ggatgaatgg ggatcatcca 164640  
tgcatgatg ggagaaagtc ttcaaaagaa agtgaagatc aattctaaca agacgataac 164700  
taaaacagca tagatgttca gagaactgca agaagcctca ttgcttaacc caaaagnatg 164760  
agagagaaga ggcaatgatt ggtagaaatt agatcttcaa ggatcttgta ttagcacaga 164820  
aaaagctgaa gatgtatttt atcacgaata gagtaccaac aaaataattt tcatagatga 164880  
gcagcatttt taccttattt atatgagaat tcttctggc cagagtgtgg catatcactt 164940  
gggaacagga agaataattt attgccttta gtgaggcaat gctgaggcta ttaattattt 165000  
gccttagtg aggaatgct gaggtatata ccaaagttca ggctagaaat gaaacccaaa 165060  
tgaagaaggc agaagtaggt ttcttgaggt gaaaacagag tcaggctctc cttaatacat 165120  
aggctcttac gactgggaga taattaaatt tagtggatgg ggaaatgaga agtgggggag 165180  
catgacctgg ttttctgact tagtgtcata gtgtgtggaa atgcattaag agaaacagaa 165240  
atcgtaacct aaggaaaatg cttgtagagg aaaagagtc ttttaaatat gataaatttg 165300  
atgtgtacga gacatttcgg tgggcaggtc taacaggtaa cttacaatac atgtgtgacg 165360  
agccctgtag aagaatctag gcattcagtg tctgctccag ggtcatcagt atgcacgtgg 165420  
cagtcaaaac aataagagtg gctgagatca tccagggaga gtgttttgaa tgagaagtaa 165480  
gcaaatagaa gacatttggg tgcgtcccaa tatttaatga atgaattaaa gatacagagc 165540  
ctgtcaagga gctgaagaat gaaaggtcag agaaatagga gcatatttga ggagaagaga 165600  
gtgatcagcg gtaaaattga atgaacagac accaagaaaag ccaagccag ccatgtagaa 165660  
gagcttcaaa ttagaaattt ttaagaggc attcttaaat gaagactgtc ctttattttg 165720  
ttttgaaaga tgatttattg attctttaat cactcaatat tgcaactggg atttcttcag 165780  
aggctctcaa gtctcctcta ttttcacaga tcaaaattct cagggtgcaga gtatgtcagg 165840  
acctattcaa gggtacacat ttggctgaca actgtgtcaa gactaaaatt tcatgatcgt 165900  
tatttcagtt catctgaaat aaatatcctc caaacataaa ccataaactt tatgttttaa 165960  
gagcacatc acctctttac atgaagtcct caagctttgt aactcctcca ttcttatcag 166020  
gtatcacttc ttcagacatc cctctcttca gtcattcctc tctttttctc tcagcctgta 166080  
tatcataaat acctatcatt tctctcttgt tgtataattg actccatcat cttttgttct 166140  
tcatttctac ctggccctaa aaactacaca tcagatcaat cctatgttct ctctttttag 166200  
gctccatcct tgtattgaac acaatgctgt aattattgtga attggtctaa tttgaattct 166260  
tgccatctaa atcttctctg agccacaggt cgtgtagaga aaatatgttt ctattctggg 166320  
tcagcatttg acgtgtttt tctgtagggt tatacattcc ctgattatta aaagtttatt 166380  
tccaagctct ctttctctgt gtattcagaa tatacctcac atgtgtcttc acagagacag 166440  
tcaatactat cagctagtag ggcatacaac tgtggccctg tcatgcctaa atcatatcaa 166500  
tttacatcct cccttctct tccactttat tccccctttc tttgtacaat gtcaatctgt 166560



ttcaaccttg	gctactgctt	cagacatctt	gcattgtcag	tcattatcct	tctcttgtgg	166620
atctccaaca	tcttacctct	gccaaatctt	ttcccttcat	cagtaaccat	aatcagactg	166680
tattttttaga	tgatgtaaat	ttctataaatt	ttatgttagc	tctggcttat	ccctctttct	166740
tcctactcct	cctcttatag	tcgtgtattt	aggtcacaag	tacactctta	aacaataact	166800
gttgccagca	gaaaacaata	tattaacagg	caattatcaa	gtatgaactc	cgaatttcca	166860
acagaatatc	cttgatactc	tgtgtatgtg	tgcttgtgtg	tgctgtgcgtc	tctctctgtg	166920
tgtgtgtgtg	tgtgtgtgtg	tgtgtgaaga	caagtttatg	acaagaagga	gagactcaca	166980
gaaggtcatc	aaaagttaga	aatttaaattg	agaatacgcg	tgaaagtaat	tttctagcct	167040
gcacctgcc	aagtgttttc	tcatagttca	tgtgaagaaa	taccaatagg	ttgttgtcgt	167100
tagtccacta	cacagtgttc	acctttaatg	taggtggcat	gggcctaaac	aaggatggaa	167160
actgatatta	acagagtgtt	atacagcagg	tgctgtacta	agggctttat	gagggaaggg	167220
taagttagtc	atataataat	atacttaaag	ctttaatttc	atgggtgggat	agttcactct	167280
catgccatca	taattgttgt	acaatcctga	attaaagaac	ttttcaagat	gttattaatg	167340
aaaaatcttt	aaagatagat	ttctatctga	ccctccaggt	caaaaggctt	ctactcaaat	167400
ttggctaaac	cctattagtgt	atgtgggttt	ctgactgctg	ccgtaacatt	tatttggaga	167460
taagatgtgc	tatatattac	agatggactt	tgggaagtga	ctgagggtga	cagaaattaa	167520
ttaaccacaa	gcctccaaat	gaaattttta	tgctgggtgag	taagaaatac	aaatttctga	167580
aacacagcgg	tatgaaataa	tgcagtatca	atggtttaagg	aagcttccta	ggttttgaag	167640
ttctatatatta	aaagatagat	atgtaacaaa	cctgtcacgtt	gtgcacatgt	accctaaaaac	167700
ttaaagtata	acaataataa	aactaaaaaa	aaaaacaaaa	aaactttgat	aagagagtac	167760
atcttattaa	tgatttcaca	atgatgccta	atataattta	gtgtcttggg	aatttaggta	167820
atgaataaat	atgggatttc	ttttcaaaaa	atataagtaa	tttttaatga	aaataatatg	167880
cagaaataaaa	tacggggtaa	agtttaataa	ttaacaaata	ttccagaagg	ccagtagtct	167940
gcgttgggtca	gaatgaaact	ttcacagagg	ggagagtata	agtgaagcag	ataaagtac	168000
gaaataaattg	atatcagtg	ctctaaagta	agaccgagga	atttgaattt	tgtaggcagt	168060
gaggagcctc	tgaaatcctg	agaatggatt	caacataagc	aaaatcgtat	tttggagca	168120
gccagctggc	atgcatgggt	ggttggcgag	aatgtctccag	tgagcaggtc	attgcagggtg	168180
tttgtctcat	gtactagttt	aatgtgcggg	cagggggggt	gctgtgaggg	aatttaagaa	168240
gcctgaatca	gcagttctta	gcaactacat	gaggggatta	agacatgtaa	aagaaagtca	168300
actgtgactc	actgaatttc	ctgaataagt	tatatcttat	tccaacgcta	ttttgaatat	168360
ttccagggtg	tatatcttga	atatattctt	gatataattca	agattcaacc	tgagtaagtt	168420
acatagacct	gatgacattc	agtattcgag	gtcagataac	aatgtaactg	aattctcctg	168480
taaagagcaa	aaatatgttt	taattatttt	acaagtactc	aacatattaa	ggaacttatc	168540
tagagtttcc	tgaaactca	agtcacgggt	gttgttgatt	ttattaatta	ggaagtttat	168600
atgaaccaat	atagaactac	actgctttat	atgtataata	ggtcctattt	ttgaaatggc	168660
tgatacagct	ttttaagtgg	tttttaaat	gagacaactt	gacatctagc	ttttaatat	168720
ttaaaggaatg	aactaggata	gctattcagg	tcatttgaat	attttgtgca	aattctcaca	168780
ctcattatgt	catgtctaca	ttcttgcgtg	ttcatataac	tcaaccctcc	ttttgaatat	168840
catgattttaa	gcactcataa	aatattacta	agaagcctt	gaggctctaa	gtaacgcaat	168900
aattagccct	cctctaatag	cccaatagaa	tacataaata	attttctccc	actatttctt	168960
ccccagatat	tgaracctca	atccgagttt	taaaaactga	cctatactag	gtactagaga	169020
tctctattgg	agatgctaatt	agattgtaaa	ctgcatctag	aaaacattaa	accagtattc	169080
taaatgggtt	gagagtotta	cttttattct	gggtcagcca	acttttagatt	cttttgtgaa	169140
tgcatctttc	gtctgcactt	ctgctctcca	aataatcccc	tgtaacagtt	atttttttta	169200
atctttcaca	cttttattct	acatctaaaa	ccttgacagg	atttgttaatt	ttatcttaaa	169260
aattctcatt	tgtaatattt	atcaaaataa	attagtttcg	agtaactata	tcttgatgtt	169320
ttttaataca	ggtttctaac	atgagactga	aatctaaaat	attatcatca	tccaaaataa	169380
tctgaaaaac	tgaaatacaa	tataatcgca	catataatat	agttacagt	aataaaatct	169440
agtatttgat	tgcgtaacag	gtgacaacag	tcgacagtaa	tatattgtac	attttaatag	169500
aactgaaaga	gtataattgg	attgtttgta	atgacaaagca	aaaataaatg	tttgtgatgt	169560
ttgtgatgat	ggatatatgc	ccccatacaa	agtctgtgct	ctcatgcgtg	ttatcccagc	169620
attttcacag	gcagaggacg	gtggatcact	tgaggtcagg	agctcgagac	cagtttggcc	169680
aacatggtga	aaccggtctg	tactaaaaat	acaaaaatta	gcagggcatg	gtggcaggca	169740
cctgtaatcc	cagctgttcg	ggaggctgag	gcaggagaat	cacttgaaac	caggaggtgg	169800
agattgcagt	aggtggatat	tgagccactg	cactccagcc	tgggogacag	agcgggactc	169860
caccttaaaa	atatatgtat	acaaatatat	atacatatat	attatacaca	catataaatg	169920
tatgtatgta	tttatacata	taaatacata	tatttacaca	cacagagggt	cattatgtaa	169980
caacatgtct	tcaaaactga	taatttttagc	catcttgtca	tataaagcct	ctaaccacac	170040
agatttaaaa	tttcttctctg	ggaagttaaat	attaatttaag	gaaatcccaa	cttcttttac	170100
aaatcaaaac	ccatgttctc	cataaattaa	catcataaaa	cattttttgt	tcatgtcaag	170160
cggtcagggtg	tctttgggga	gtttagaatt	acctgatacc	tctgttcttt	tatatcaact	170220



```

ccataaggaa atcccactgc attacttggg atgtgtttct gtaggcctct ttagtgtaaa 170280
atttaaaaga aggggggaagt ttaagtaata tcaggattaa tgaaataatt ttcaagtaaa 170340
atataatttcc atgggtttac acaaagtatc ttcactgtgc agtgactctt aacagcagat 170400
gcgtattttct cactcccatg aaaagttctc cactgggtcca ctgactgtga tctttgtgtc 170460
ttttaccctt ttggatccag gtggaaggag tcccctcttt gtgacactga agtttcccta 170520
gaaaaaagta ggagtcatat gatggttctt aaagtaatag acatcctttc tgctcacatt 170580
tccttggtcca atgcagggtca catggtcaag cctgatgtca tcagggtggg aggtgtgatc 170640
ttcctatagg acaggggagc aaatactttc aatctgtcat aaaacctgaa tgaaatagta 170700
agatatctgt taaatctatg tttcttaaac acacatgact gtctacacag gggtcccttg 170760
tgcattattt gtaatgttaa tacataaaac atgaaggaga gtgtgtctcr ttctcatgat 170820
caaattttac ctgttatctt ttctctatta tcaaaatgct ggtagtcttc taatgatctt 170880
tatttgttta tttatttttt agttattgaa taatgataat ggtgtctatt tcagatctac 170940
aactttctga gttttaaagt catggaggaa aaactataat gaaataaaat ataattttat 171000
ctaaaaataa ataaatttgg tgttaaatcc aagcatattg ttattaacag gtaatacttt 171060
tattgatttg ggtgttttag gagttttagg agcatagaaa tatatcatac ctgtgtttac 171120
tttattaaag aatcttttgt tagaacttcc cagacagcat ttttgagat ggtaacctga 171180
gatcatgggtg tcatggtgtc cttactaatc acttttaata ggatatttaa accccctttc 171240
cagaaagaaa acaaggagct tccaacttcc catgtacccc atcacaaatc atccttcttc 171300
tggtactgtg taacacatct acccgtcca acacatctac cccatcctct gaggtcattg 171360
tgtaataatg taaccaactt ctctcatttt tctgcagttt ttctccattc ctgtgaattg 171420
ggtgtattag tcccttctca tactgttata aattaatacc tgaaactggg taattcataa 171480
agaaaggag ttaattggc ttacgtttct gcaggctgta caggaagcat ggttggggcg 171540
gcctcaggaa acttatgatc atggcagaag gtgaagggaa agcagacatg tcttacatag 171600
tgggagcacg aagggtgcgt gaggtgccac acacatttaa acaaccaggt ctctgagaa 171660
ctctatcaca agacaacact aaggagatgg tgctaaccat tagaaaccac ccccatgatc 171720
caatcacttc ccaccaggcc tcatctccaa cactgggat tacaattcaa cgtgatctct 171780
gtccctacac acagccaaac catatcatat gtcaatgttg ttctctccct gaaaacaaga 171840
acaaggctgg gcatggtggc tcacgcctgt aatcccagca ctttgggagg ccaaggcagg 171900
tggtacagga ggtcaggagt tcaagatcag cctgatcaac atggtgaaac cccgtctcta 171960
ctaaaaatac aaaattagcc aggcattggt gcacgtacct gtaatccag ctactcagga 172020
ggctgaggga agagaatcac ttgaacctgg gaggcggaga ttgcagtga ctgagatcac 172080
actactgcat tccagcctag gtggcagagt gagactctgt ctttaaaaaa aaaaaaaa 172140
aaaaaaaaann gagcaaaacc ttcccctgat gctctatate catacatgtt tttggctctg 172200
tggaaccatat ggactctatc acaattactc aactatgtga ttgcaatgaa aacagctatg 172260
acaatataaa aatgagtggg cataactttt ttcaataaaa ctttattcac agaaacaggt 172320
ggcccaatgg tggcggtag attgtgact cctctgctat agtctttctg ttaacctacc 172380
ccttactgtt tcttcaaaac atctgcttaa aaaaaggttc tatttctcat ggtttccaga 172440
ttcttacctc tgtttcatgc ttgaaaactc tgcaatttga cctttatacc caaagttttc 172500
tgacagcttg gggcaagagt ggctttctga ctgagtctga agcattatct gtgaattatc 172560
ttgggagtac atttccgttt ggctctctcc acttttagact tcttaccoca tggatactct 172620
cctctgttcg gtgaaactat ccacccttg ttctcttcca gtatgtgagg tttcttctca 172680
gcctcttcct cgttaagaat gcatttaact cttaggatt gcaaaaacgca acaataacca 172740
aaataaacia gaattttttt cagatttcaa acctaacatc tggagttaag gagttcaagg 172800
ttggtgcagc cctttggcaa tgtggccaga ggtotaagca ttttctatt tctgctccgt 172860
cttccatatac attttggctt ttgttctcat gtttgttgcc ttatgggaac ctgaggtttg 172920
aaatgccatg gacaatgctt gtgtgtctaa gacaaaaaaa aaaggacaaa attagtttgt 172980
tttttaaaatt ctttattcaa tgaatcagaa aggaaaagtg tccctaagac cctggcagcc 173040
atgtctcctt aggtatcctt ggctgaatt agccagacc tgaactttgt cagttcagtg 173100
catgagagaa gctgggaagg ttcatatttc tagccagggg cactgcctgc ttctccaaat 173160
tcggttttca ttaggaagaa agatgagtta aatggatatt ggctaaacaa ttgacatgtc 173220
tcccagagtg acctgtgcag gatattttct ccccataac tcggtgttag tgttctagga 173280
atttatcttt cgactctttt ttctgtttaa ctgcattgcaa ttctgtgca gtctcattca 173340
ctatcttagc atgacacrt aagggtctgat gctcaaatct ttctaactca gattttagta 173400
tccatatacc atctcctaag gggtttcacc ttctgttttc cccacgtgta tttcaataa 173460
aagatagctc gaaagtaatg ctgtttcttt aacgggtat ttctcataca tctctacctt 173520
gattgtaact ataaaaatga taatcaagaa actgaaactc aacttctcct ctactcctt 173580
caactgctat actcaatcaa ctgccaagtg ccatttacaa agtttcttaa catctcttga 173640
ggctattaac gtattttcat acccactgtc attctcatga ctagtttctt ggtaaaacc 173700
tgtattattt ctactctgt ctattgaaga ttacattctc acctgggtat tctgcttccc 173760
attaatactg ctatcagaac actctctctg aaatagaaac catttttcta ggttaaagcc 173820
cttcaacggc ttgtcataat taacagcagg taactcatac agtccctaaa caatcttatt 173880

```

tatcccaggg	tattgctgag	gactttcttaa	tctttcactc	agttcagact	cactottaca	173940
gctttatttg	gtttacaaag	gatattttacc	tgatatctac	caggtatctc	aagctgcaag	174000
actgaaatca	ttaccttcaa	ataacctcca	tatctatact	ttgtcccttg	tgttcccttg	174060
tttagaagat	gttttgtcaa	caactccatt	cctcctagaa	gtttttagt	aatcatccct	174120
gattttcttt	ctcctctgct	tcactactaa	ccagcctcct	tttttctgat	cagccacaac	174180
atcctattgc	ttcaattctc	ttaatgcctc	tcaaataatac	tcactccgct	ctagtgtggg	174240
gattacagcc	ttacttccag	ctactgtaat	tacttgccta	aaatatgacc	agattctttt	174300
aatcttttcc	atctgttcgt	taaaagaatt	tattctagag	gttttatttt	ctacaatgga	174360
gctacgatta	tattattcac	ttcttcaa	aattcaatat	cttaccatta	ccctctaaat	174420
gaagtccaaa	cttcttacca	tagtttagaa	agcttttcca	gagctggaca	catctatcat	174480
agataatgct	gcaggcaggt	tgtatctgtg	aacctgtcat	cacctccctc	tttgtactta	174540
ccacactgag	gatgctgggc	actgcatggc	ttcccttatt	accatcatca	tggtccctac	174600
tgtaaactga	attggaagat	tacatgtcct	aatcctccac	cctttcatca	catcatgggc	174660
cttctgatct	caacctgat	gttcttcccta	tgccctggctc	tgtgtttggc	aaatggtaag	174720
ggccagtctg	tgcccttcca	tgcattggatg	actgattggg	atgtcacctt	tgctcctatc	174780
tctcatccca	tcactcccca	tcgctgcaact	tgactcacta	gcgttttctc	ttccctaaac	174840
agcagattca	gtttgggtct	tgaggccttt	gcactttgtc	ttctctctgt	gtgcaagcct	174900
ctgcttttgt	tacactgctt	gaaggattgt	cacctttcaa	aacaggagac	ttccgcatgt	174960
ctgagaaaaga	gttttaaggg	ccgttgcatg	ttttggctgt	ctcccttttc	ccttttgcc	175020
gagaacgacg	tttcaaatag	aagctgtctct	ttcaatctgg	tgaacggca	tgaggagaaa	175080
aaccatattt	aacctcagag	tgaatgcaa	tgctcagtga	aaaaaagcta	ccgtgagcca	175140
ctgagatttg	gggactattc	cacagcacac	tccaacctag	ggaaatctga	caaaaacaga	175200
ccatcatatt	tgaagatttg	aattcaactt	ctctctggtc	tcttagcctt	ctatcaatgt	175260
atgctagcag	agaatgctac	cttcttccca	cagtaagaaa	agaagaagaa	atatcctacc	175320
ttctcctcag	atgtaaatta	tgagcttggg	tttccctcct	gtataagttc	tatcagcttc	175380
aatgcatgca	taggtctctt	caaaaaataa	aataaataaa	atagctaaag	atccttcccta	175440
aatgctactt	atcactccat	ttaccatgtc	cctagtgtga	aacaactgtt	tataatcagt	175500
ttaaaaaat	agttttgttc	tctccattcc	cattttgtga	gctccctcca	gcttctcaac	175560
ctttttcatt	tgtatttcca	tcctggcaat	gcattgtaaac	accctttaat	gttattagtc	175620
gtatgtgtgt	ccctaaaggc	aacagatacc	ctttccgttc	ccaattttatc	agaagactca	175680
tgatatttca	actttctgga	ccattccaga	ttttataaaa	tgtgaccatc	atttgacttc	175740
agtaactaca	attgcatcat	tttgcctctg	ttgctttgcc	tctgcttcag	attgatctct	175800
accctaaaat	gttgctcctt	cacaaattaa	catataactg	aatatgcaac	attttcaact	175860
gggtatctca	agatacatct	caaaactaac	atccccaaaa	caaaatatca	taatttttgc	175920
ctttacaaat	aaaaatctgt	cactacaagt	tacgtttcag	ttttagaaaa	tcataagaaa	175980
ttgaatttta	gaatatgcat	tctcttgtgt	cgggttttga	aagacacaa	gttttgtgtt	176040
ttagtgtaat	atttttaaga	tttgtccaca	ttgcatcaga	tttgcccttt	tttattgctg	176100
agtcataaatt	tgtatacaga	tataccaact	gaatccagtt	atataaaatt	catctacaat	176160
tattttgtgg	acttatgtat	ttatttttct	tagggaaatg	tttaacagtg	gggttttctg	176220
gtcatgtggg	atgcataaat	tgaatacga	aatgcagaac	tttctctaag	cgactattgt	176280
taggtgtctga	gattttaacc	tacttataag	ctaactgttt	agttttattaa	tgcccttatgg	176340
atgcaagcag	aagacacaaa	acttctgggt	cagagaaaca	ggactttatt	atccacagta	176400
gagcaatcag	catgagcatc	atgtttgact	tggttttttg	tatcgtccac	gtcccacagg	176460
gatgacacag	agggcccaga	tggttggtctg	aagattacag	gagagggaaga	gtgatcttgg	176520
ggaactcttc	attttaagat	aactggtgag	caaacctaaa	ctttgtcttg	gataaaaaa	176580
tttctcgttc	cctcaaagtg	gtcatttgca	atcataaccc	tgagaaatgg	cccaggttaa	176640
aaaataataa	aaagtaggag	agtttgatac	tagatacttt	tctagatatt	ttgaagtatc	176700
cagaaaacgt	gtggtattag	gagagaccca	tgaggagcga	tttcacaaca	gctgagtaat	176760
tttacattcc	cattagcaac	ataggaaaat	tatttaattt	cagtaattct	agtaggtgta	176820
tattaacttg	aatttccttg	aagaataatg	gtactgaaca	tcttttcttt	tcttgcttat	176880
tgagcattca	tatgtccttt	tttgcaatgt	gtatgttcag	aattgtttct	ttttaaattg	176940
ggttacatct	tcttattttt	cagttttaag	ggttctttac	atattattaa	tacactttct	177000
ttgtcagata	tgtgcactgc	aaatattttt	ctctcagttc	gtggcttaca	ttttcatggt	177060
cttagagaaa	ccttttgaag	aacaaaattt	tatcttggtg	aaatgtaatc	ttttattttt	177120
tagcacttag	tgcttttgta	ttctaagaaa	tgtttcccca	ccccacaagt	ctgaaatttt	177180
tatcttatcc	ttctttttac	aagtttataa	tatgattttt	gcttttttgt	ctatatattag	177240
ttttgagtta	tttttggtga	tggtgtgagg	tggtgtgatt	taccttttcc	aggtggaatc	177300
aactgtttct	accaacattt	atttaaaaga	ttgtcctatt	cctgtttctt	tctgattgaa	177360
ttatgtttgt	cctttctccc	actgaattct	gatattttct	aaaaatcaaa	taataatatt	177420
tgtgcatttc	cagttttggg	ctctcaataa	tgttctttct	ttatgccaat	atcacactgt	177480
cttgagtgtc	gaatatttaa	attaagtact	ctagccagag	gtatttgatg	ggttccaatc	177540

cattacattt	attattatta	ttataactat	aataataatt	attaatgggt	gtttacagaa	177600
ggtaggctag	tcctgtagca	accattttgc	cataactaat	agtgtgatat	tctgtttact	177660
tacaatttga	tatttttctt	ccttttagtaa	attggagatt	gccagaaatt	ccttactatt	177720
tctcccatg	aaatagggaa	gcaaattcac	ctcaccttga	atgtgggctc	attttattga	177780
cttcttgtcc	aacagaaggt	gggacaaaaa	aaaaaacatt	atgttattgg	acttccaaga	177840
ctaggtcata	tgaatccttg	aagcttcttt	ccagctctct	tggaaacactt	tctcagttag	177900
acccgagatg	ctatgtaagt	ctaacaacca	cctcactgct	gtactgggga	agacacatgg	177960
agggtgtctg	atggacagtt	cctgcctagt	ccagaattcc	agatttcctg	ccaagatgct	178020
aagaaaagtg	gaatagccct	cttggatctt	ccagaccagt	tcactctgcc	ttagtgttag	178080
tgtgaattct	agaaatacca	tgttgaaagg	aaagctccat	gggtaaatac	gcttctgcta	178140
ataagtggga	taccagaacg	gctgttagca	gctgtcaggt	ggaagtgggt	ctcgtgtcag	178200
attccagtg	aagttttcag	tgtctctttt	ccttataggc	ttaggatgtt	tctctactgc	178260
tcccagcaga	ggtgctcagc	atactccaaa	gaactctgac	tttgatatcg	tgccaatcag	178320
ggttcgaatt	cacaaatttc	catttataat	ctgtataaca	ttaaatcctg	ttccaatggt	178380
tctgtcattc	agctttgtca	tctctccata	cagaaaaata	cccacccaat	gtcatcccac	178440
ctctgctgcc	taaacactgt	attaatgtta	agtggcttcc	ccatgctatg	cactgtctat	178500
aaaatagggg	tagtatcacc	ttcctctgga	tgcattgtga	aaagtcaatg	tgaccatcca	178560
cgtaaaaaat	ttgatgttag	gcacagagga	tttgcttaat	aaatgttagg	tattattcac	178620
atgaggactt	gatgaaaaaa	aggtaaaagc	atgtatttat	cacataaaaag	ggtcaataaa	178680
atgccagttt	ccttcattct	ttccttttta	gggcactttg	aaaatatgga	aaactagtta	178740
caacacagcc	tcaagaagct	catgattcac	tggttactta	tacaagtgat	cctaattatg	178800
gcttgagttg	tagttgtttt	acaaaatgta	aaaattattg	agataataat	tagttattta	178860
caaaaacaatt	ttgctcgtag	agattttaaa	ttctgtcatt	aaaaagaccc	acaagcctct	178920
caattgcaaa	gaagttagca	tcaccttttt	ttttttacca	aattcgaaaa	aagcaataat	178980
ctaactacag	aataaaccat	ggcggtaca	gtacaaaaaa	tgttcttatg	tttttaaaaag	179040
ttgtgcagtc	cattgatcag	gcaggcagtg	gtggggagaa	caaagccaaa	cactgaatac	179100
acgttttgc	acaaagaacc	tagtctaate	agttccattt	agaaatgcac	tttcatttga	179160
ctatatttgc	aaatcatcat	ttttttttct	ctaaaaataa	tatggagtgt	ggaaatgctt	179220
tcctatgcag	aaaggttttt	ttattttatt	tgctgagttt	caggaaatga	agcttagttg	179280
ctaccttgca	tagaaacacc	aaactataga	atgcacaaat	ttagcacaga	gtctgccacg	179340
tatggcaatt	ttgataccgc	tacctaacaat	tcacatttac	tgcatccatc	caattacacc	179400
catggtatca	atgcacagca	cttagagcag	gctggcagtt	cacaagagtg	caacggatgt	179460
cacgcagtc	gcatacagt	caatctgctt	cctgatatgc	tttaagaaaag	cctgttccctc	179520
agactcacct	gaaatgcaaa	gagagaggtg	gaggacgatg	gcccattgtt	cctcttgtgt	179580
acctacactg	cagttcttgt	tgatgaccat	ctaattgcaa	ctgtcccaca	aatctttcaa	179640
agagagttaa	ggtgaagcca	gtccctatgt	taacaaacct	caaccttatt	tcagcagatg	179700
gtaaaacttg	attttgttta	agaaccaaag	cttcaacagc	ttcaggaaaa	catgtgtttg	179760
taattcatct	ccattttaat	gccacccttg	catggacgag	ctacatctcc	tgctccaggt	179820
gaatgggttc	ttacgttctg	gacactgtgt	cttcatttca	gttattgata	cctagagttg	179880
gtgagtttca	agccaatctc	taagtcaaga	gtttatttgg	gagttgggtt	tccttttgaa	179940
atatactccc	tggagcctca	ggctgttgcc	tgttcacctg	tttcttttcc	agaccaaate	180000
atggaaagga	ccaggttatc	tctgatccat	tgtgggtgtg	taaatatcaa	gaaggataag	180060
gtacctttcc	tggacttgag	aacttatagc	ctaataagag	aagacaggaa	taaaaaaaga	180120
aaagttttag	ttgctatgag	actggcatat	aaaagtacag	ggacaggaca	aagaaaaaaa	180180
tggaaaatta	tcagaaatgg	gtataaagat	ggcttcatag	aggaggtaaa	ctctgtctga	180240
gaaattagaa	aatgccccaga	aagacatcaa	agagaaggat	aggctgaaag	gtggaaggca	180300
taaatgctgt	taaaaaaggc	aaagagtcac	gaaatagaca	atgaatttct	agaaactaga	180360
aaactggaca	caatgggaca	gggcatggaa	tacaattaag	tgacctctga	tttaatgttc	180420
aagctagtat	caaaggagat	gctgtgaatt	ctaactttga	aatagcacac	ataaaccaga	180480
gctgccccag	gcaaattgtg	acatatggtc	tccttaggca	taaagttaat	ggtatagtca	180540
gaaggttaat	aacgtaggaa	atgaattgaa	atggctgctc	tgacacacgc	tacaaagttg	180600
gaacctgatg	caatgtggta	ggcagaattt	tggtacacta	gactttcacc	cccgggtgcca	180660
gctttgtgat	tatgttatat	gcattggcaga	agggaacttg	agttgttaatt	aatgttgcta	180720
ttcacctgac	tttaacagag	ggaggtgatc	ctgggtggtc	caggtgggccc	cagtgtaatc	180780
acacgagtc	gtactagcag	aagcagaagg	cagacagatn	agggaaagag	gcaggacaga	180840
gatatttgaa	gcctgacaag	gacttgacat	ggtatttaata	gttttgaaaga	tggcggggaa	180900
cataaggaag	ggggtacagg	tggcctctaa	aagctgatga	tgacctctgg	ccaacagctg	180960
ttgacctgga	gaagaatgtg	tgttaaacaa	gtgtggatgg	tgatgataat	aatgaagatg	181020
attttgtcta	tatgctgaga	agacatcatg	ttaggagtga	ttacagataa	tgacgcaaaa	181080
aaggagcaaa	gtgttgaggt	agaggtgaga	gaaggggtga	gctcaagagg	caggcaggaa	181140
tccttagtca	ggcacagtgg	ctatacatc	cttttcattt	ctccttggac	ttgcagatgc	181200

gaaacagatt cctttttctca gatgcccatt cattaggggt ggccatgtga ctgaattcca 181260  
 gccataaaga tcacatgtgg cattttaagg cagaagcttt tcccttctcc cagcagaatg 181320  
 gaaaagctga agcaaatagga ggagatgcac catacaatcg aaggaagcag agaccctcaa 181380  
 tatggaacaa cgagaaacat ccctgcagta atagttgaac aaaaacacac tccattgtg 181440  
 ttgtgccctt tcttgacatt ttagactgtt cattttcaca gctggaccac tgcaaccatc 181500  
 acacatgtga gaagaagggt ggtctagaag gaagaaagt agggaaagcta caaatattca 181560  
 tgtttgttag tggggaccat gaaggcgaga tagatttcaa tttcctcgge cgggcatggt 181620  
 ggctcatgcc tgtaattcca gcactttggg aggccaaggc gggcaaatca cctgaggta 181680  
 ggagttcaag accagcctgg tctaactgg tgaaaccca tctcaactaa aaatacaaaa 181740  
 attagccaag tgtggtgatg ggcgctgta accccagcta cttgggaggc tgaagcagga 181800  
 gaatctcttg aacctgggaa gcggaggttg cattgagcca agatcatgtc actacactcc 181860  
 aggtcgggtg acagagagag actccatctc aagaaaaaaa aaaaaaagat ttctattttc 181920  
 tccataaggt agtggaatg ggtctgctg acctctgag tgtgtaggag tggagggtga 181980  
 gaatgggata cccaaaagca gaggagctca tggaggacaa gaactgtgag gactggaaga 182040  
 gaaagcatgc taggggtagg attgcagaat agagctcaaa cttccaggtc aatgatgagg 182100  
 gttcagcagg acttgacac tgggcatgtg cagtactcc atgtctcaat attgtctgta 182160  
 tagtgaacag atgttttgaa tgtgggagac tagaataaat gcatttttat aaagattttt 182220  
 gattaaggca ttagtaataa gcttcagtag aactgtttac tttttctac tgcaaagata 182280  
 tcccaagggt tctagcctt catctcagt tgccttctta atagttacca tgaatatag 182340  
 taaggccatc cataatttca cacctatact tatcatctca aggcaagcct tgttagacag 182400  
 gcataagctc ataataaatt acctgatcca gataaggctt tttttttgt tttcccaatt 182460  
 acactcatca accacaaatt aagattaatt acatagatta cttgggtctgt aaacatttaa 182520  
 ctaccatcaa tggatgaat gaccatgtta aaggaatttc ctaagtcaa tagcagagta 182580  
 ctaccattcc ttgagtatgg atagtcaata ttttgtaaag ctataaatac gggaaatctgt 182640  
 aaagaatata tctctgatgt ggagtcata ccaaatctgt tcaaaatgcc tatgttaaat 182700  
 gtaattagta tattctgatt tataatctac catataaact cattaanaaga aaccacaaag 182760  
 cagctcgcca ttaaaaaaaa aaaactgtta aattgatgga aatctggtac ttatttttta 182820  
 ggcatatgc agtgtgcttc attattatac acagattctt ctactatag tatcgtttct 182880  
 gttaaaagct gtgtgtgta ctcatacagg atacttacag agcgttttta aagcagtaca 182940  
 taaatcatag aaatttttc aaatgtgagt taaattgcaa gtcacaattt agttaagttag 183000  
 attagtgaat agaaataagt gatttatttt aatagccaaa aaacaagctt aaagcagggtg 183060  
 aaaggcatag agcagacaaa ggctaaattt tatggagcat ttgcacattt gcaagatttc 183120  
 attaaatttt ctcaacacaa ttttctctt ctagtaatat gggagactag atattctaac 183180  
 tggctcttct actaattaaa aacacctaaa aagtcttggg aaatttgttt tataaatgaa 183240  
 tcaatgacag agaaagaaag taagaattct catttgaaa ctaaaatagag aaataaacct 183300  
 tgagaagaaa gaagaattct gaaacagggt tgaatttgag gatatacccc aaaaatcttg 183360  
 acttgcaaat ttgtttgcac agcctctcag agtgcaggaa acaggaggac atccagggtcc 183420  
 ctaaaaagat gtgagttgat cataagtact cttgatgagc tggaaatgatg agaaggctac 183480  
 actcttagca tagctgtgag ccgaaactaa agacaactcc tttctccagc tgatacaag 183540  
 taaattgtcg ataaacttga cctgaatgg aaggtaaaca cgtaaatata cagagatata 183600  
 aacatacaca caaacctccc atcaaccctt taataagaaa gaaaggaaaa agctagacct 183660  
 cctccaatt tacacttaac gcatggtaga tacagatata gacatggata tatatcaca 183720  
 aaaaccaga gtttatttta taggcaattt acatacggca aaaagaatac aaatcttttt 183780  
 ggagaatttc agcttcatta ctaaactcat ctctagagaa tgttactaga aaaatgagca 183840  
 ctaccagca gatttcaata aagaatcaag aaagaaagaa ccatgaggaa gaacttgcaa 183900  
 acacaagagg catcagaag aaggctgttt agtaagaacc cgtaggaaatg taatcacaca 183960  
 tggaaatata aatctatatt taatatcttt aaaggaatga aagactagt ttaaaatatg 184020  
 atcatggaaa aagaaaacat aagaataaac atgatgattt ttaaaagaag aaccaagttag 184080  
 aactttcaga tataaaagaa ccagtgtgctc tgggtatgac cacagcaagg acagagctga 184140  
 gtgccttagt gaactgagg acctgaataa attaccggga aggcagtga gacaagcaag 184200  
 gaataaagt caatgaagga caacctagca caggtggaag aaagaatgtg aaggatgaac 184260  
 atatctctca ctggagtctt gaaaacagaa gtggcaggaa aatataaaga ttgtaactga 184320  
 ggattattga tgaactgat atgatgacaa ttaatacatt catgaaaaca tataaatact 184380  
 aaccaggata gttaaattaa aataaaacac tcacacttta acacacctta tgataaaacg 184440  
 ggagaacacc aaagtaaaaa gagcattgtta aaaatagtca gcaagaaaag acagattaaa 184500  
 tcccagagag tgacagttat atcaactgct aactttccaa gagcaaaaag gacattagca 184560  
 gaaagttagg tgaattttgt tctgtattgt tctgaaagaa aataactgcc aacctacaac 184620  
 actattgcct gtaaaactta caagaggcgg ggcgggcgg ctcaagcctg taatccagc 184680  
 actttgggag gccgaggcgg gcggatcacg aggtcaggag atcgagacca tctgggctaa 184740  
 cacagtgaac ccccgctctc actaaaaaac acaaaaaaat tagccgggagc tgggtggcagg 184800  
 cgctgtagt cccagctacg cgggaggctg aggcaggaga atggcgtgaa cccgggaggc 184860

ggagcttgca gtgagccgag atcgcgccac tgcactccag cctggggcgac agagcgagac 184920  
 tccgtctcna aaaaaaaaaa aaaaaaaaaa aaaaaactta caagacagga gaaaccaga 184980  
 gattttgtcc aagtcaatcc tgattaaatt aaaccacaaa taggagtaga gaaataagaa 185040  
 cttagagagg actaaaatgg aagtggggag ttgagagcaa aaatctgttt acatgggtaa 185100  
 ataaaaatca aaataatata aaaataatca tttatcttgg gataaaacaa aactaaacaa 185160  
 gatagactaa attatgccat atcgatacaa tatagcctgg gaaaaatgtg aagataaatt 185220  
 tctaaatttc cctgtattgt ttaggggaga ttaccagttt ttaaattatt tttaaactta 185280  
 ggtaagttaa ctgtaataaa aaaattcctt gggtaaccca gtagttttta aggaactgaa 185340  
 tataatgaaa atacaacata taattcccaa aaaagtaatc acaaattaat gagaaaaata 185400  
 atcaaccatc tgaatctata agtaagaaca tcatataaag gtgggataaa tagaaataaa 185460  
 aactaatatt ctaaaaataa gttagtatg tgaataatta gaatatgtat aaattaaata 185520  
 aatactccag gtaagaaatc atcgatatac agcctccagg gtcttagaag gaaacaaaca 185580  
 gcacacttaa ctaggacaac ctggtgacag tacaatagag aaaatatttt taaaagtata 185640  
 gatataccct gttttataac aataagaaga ttctcagggg agggatataca aaatatacca 185700  
 catgcaactgc caatgaaagt gtaaaactgtg taaactttct gaaaacaata caccacaaaa 185760  
 acctgaaaaa cgtaatatccc ttaggctcta agaaaagtga cactaagaaa atatgtatct 185820  
 atagacatgc ttattattcc atgatttata attagcataa atgtagcaac tacttaaat 185880  
 tccaacaaca tgaaaatagt atttaaaaaa ttcattaata gcatggggca atactcattt 185940  
 cataaggcat atgaatcaag aataaaatcc aaagaactcc aactgattaa atagactccc 186000  
 cttggccaag agtatctcaa ggaaacccaa aaaactagtt cagaccacaa taggaagcgg 186060  
 ggtaaacaat gcctccttat accttccct ttggagttca ggcatacaac tgactgacat 186120  
 taattattaa atagagatca gaagactgac aacacagact ttttgtaaca agaagatacc 186180  
 aaattctaac ctgagtctgg tatagcatca catgacagat agcagcaggc tttgaaggaa 186240  
 atcaaggat ataatgtata tcaaatgtat ttctttgaca tattttgaaa tgaccctgaa 186300  
 aagctgtttt ttgtagagga aatttgcagc ccgtaaataa tctccttccc ttactagata 186360  
 tttctagag tctgacatcc ttttaaggat gataggaaat acatatgatc tattgtctct 186420  
 gaagcctgct acctggagac ttcattccca taacaagaac cttgggttcc ataatcccc 186480  
 ttatcttaac tcaattgttt ctttctttt ttttctttt tgaatttcag gctaagttaa 186540  
 ttttttatt ttttttatt ccaataggtt tttggggaac agctgggtga tgggtacatg 186600  
 aataagttct tttagagtg tttctgagat tttgttacac ctatcacctg agcagtgtac 186660  
 actgtaccca gtgtgtagtc ttttatccct caccaccccc tattctttcc cccaagtccc 186720  
 caaagtccaa tgtatcattc ttatgctttt gcactcctcat agcttagctc tgacacatga 186780  
 gtgagaacat atgatatttg tttggtttt tatgactgaa ttacttcaact tagaataata 186840  
 gtctcaaat ccattccaagt tgtgtgaaat gccattattt cattctttt tatggctgag 186900  
 tagtatatat gccacaaata tatatatata ctaaatatac atactatata tatatatata 186960  
 tatagtatat aaatatatct aaatatagta tataaatata taaatataat atataaatat 187020  
 atactatata ttatatatac aaacatatat acacatatag tataatatat atactatata 187080  
 tattatatat atacacacat atatatatac acaccacatt ttctttatcc acttggtcat 187140  
 tgatgggcat tccggctggt tccatatatt tgcaattgtg aattgtgctg ctagaaacat 187200  
 gcatgtgcaa gtatctttt tgtataatga cttattttct tctgggtata tacatagtag 187260  
 tgggattgct gtagtcaatg gtagttctac ttttagttgt gtaaggaatc tccataatgg 187320  
 ttgtactagt ttacattccc accaactctg ttccctttt actgcattcc cgccaacatc 187380  
 tattatgttt caatttttg gttatggcca ttcttgagg agtgaggtgg tatcgattg 187440  
 tgggttgat ttgcagttcc ctgataatta gatgttgagc atttttccgt atgctgtgg 187500  
 gccattttta tttttcttt agagaattgt ctatcctcc ccttaggctc cacttccaat 187560  
 attgggggtc acatttcaac atgaaatttg gagtttgaa acatgaagtt tggaaatata 187620  
 caaactatat aagcaggatt atatatagac aataaagttc aatgtatttt tcaaggcact 187680  
 cattatttat atatttatat atttatatta tttatatata aaaataatat attatttata 187740  
 tatataatat atatatata tatataataa tatatatata atataatata taatatataa 187800  
 tataataata ataaatatta tttatatata taaatttata tatataaatt taatatgcta 187860  
 tataaatata aatatataaa taaagtatac tccacaattg cctaatacag atttctttta 187920  
 caaatccata ggacacctat aaaattgaca gatataaaat gtgtttccat ggtattttaa 187980  
 agatttatgt tatgcagata actgcctctg aaagcagaaa cattaatgtg caaataaata 188040  
 acaaaaggac aactagttaa cagatcaatg aatgccctct taaaaaagtc ttgggtcaaa 188100  
 gaagaaccca aaatagatat tagaaaaat ttttaaaaaa tagataatga aaatatatat 188160  
 taagcctagt aacttacagt cagacagaaa tggagttta gcagatatta gaaaagaaaa 188220  
 tactaaaaatg aataagcttg gcatgtattg aggaagttat aacaagacaa ggaagaaaaa 188280  
 ccaaagacag aagtaaactc catggtgaac tgtagttccc ataatcccc caactagtgg 188340  
 gagggacctg gtgggaattg aatcatgggg gcaattacct ccttgctgtt ctcatgatag 188400  
 tgagtgaatt cttgtgggat ctgatggtt tataaccttt gctctgcact tctctttgct 188460  
 gctgccatgt gaagaaggac atttttgttt cctcctcctc catgatcgta agtttccctga 188520

```

ggcctcctca gtcattgctga actgtgagtc agttaaacct cagcctcagg aatgccttta 188580
ctagcagcgt gogaatgcac taatacatcc catatctcct taaaatgtac aaaaccaagc 188640
tttgcgctga ctaccttggg cacatgtcat caggacctcc tgaggctgtg tcatgggtac 188700
atccttaacc tttgcaaaat aaactttcta aattgattga gacttgtctc agatactttc 188760
tgttttacaa attacctttt gttctagacc actcacaaga gtgacctatgc aaaatggaag 188820
tgtgccactt actcagctag-aaacgctcct tagcttttcc tctcactggg aatgaaagtc 188880
aatattctta cottgacctg tcaggccctg tacttaagaa aagaacagca tgctttgacc 188940
taaaattacc tctgaatggg ctttgtctgt aaaatggaag agacatggag aggcgcggc 189000
accgaggtg ctttctaaaa taaggggtgc cttaaaagtc gtgttgaagt ttgtcttcc 189060
ggaataggaa gtgggtggaa tattttaatt agggaggtga tgtgattgaa tttgtttata 189120
aaattatttt tggttaaatg taatagagaa ggaatttgat acagtgggta ataatgaaga 189180
ggaacaatt tttagaatat taaaatatcc caggcttgaa tgaacgtaga gatggagaaa 189240
tgtgtaaaga ctgaagaggc tgttgagagt tagatggaaa agacaagtgg acggattagt 189300
gttgggataa gaaaggaaga aaggagtcac tgatggggcc atgtttcaga aggtttagg 189360
gtgaaggctg ctgcaattca ttgagaccag agcctttgaa agatagataa gcatgcttg 189420
gtgcagtggg tcacatgtga aatgtcctac gaaacataaa agtgaaagga ggagacaaa 189480
ataataaaaa atatatattc ttgatgtcaa aactaaaata taaatatttt ggagactttt 189540
gtgtgtattt ggatgctatt ctcattgctt ttttctgga aaaaataata caaaaacaaa 189600
attattttcac gattttctca tttttgtcct tctgttttta tcagataaat gctaataact 189660
ttcatatctt tcttctcctg cagtttccca tgtcttccaa atttatttca catcattatc 189720
ttacttttcc tcaaggccgt ttttctcct gactacagca atgcagattt caattctgcc 189780
attgaattcc cagacatatt cgtcatccc attttcatcc cccaccacc tgccattttc 189840
ttcgtgttaa cttgttttcc tgactcacag aaatcacctt ttoctgtata catttttagg 189900
atgtcagact ttattctaatt gatttctcct agttgcccc caaaattgta ttctacrgtg 189960
tgattttaaa gctgaatttt caagatgata ttcatatct atattttcac aagcttttct 190020
tctatgaatg ttattgtcag ctgtcagggt gtgagatggg acttgatact acattctttc 190080
caagctgttg cctgaatcgg ttttaagacaa agtcattact aggtgttaa ctgttgctct 190140
gcaaaattga gcagcacgta ttttaaccact catacttctt agctctccaa cacttttagt 190200
caatagagta tcacaaacacc tgtaatttcc agtgagcaca gctcccaggc tctctctctg 190260
tctctctctc tctcttttgg tttttttttt ttgacctcagc ctctcgggct ttatttggg 190320
gatttgcacc ttcaccttct gcatgcagtg tgccatggct agttccctt acatttttgt 190380
catgattttg gttagaggcaa tgatttctct gttgtagaaa aaagggcagc aacataggat 190440
tacaaaagta ttgctaagaa attgcaacaa gctacttgca gtttatgtat gattactcac 190500
agtttttggg tgggtgcaacc cttaatttct gaaccagcag gtgaaagatg agagaagaat 190560
aagagacctt tctatagatc ttcaggcaga caggcttggt ggcttcctaa tatttcgggtg 190620
ggatactggg atgggttggg ggcaacatac tgtttaaata agtagacagt tttttaaaaa 190680
taatgaattg tgtctctaag atatatattt accagaatga tactatatag gaatatgcgt 190740
aaacaccttg actttgatct caagcaaaac cttttcccta ttatccctcc ctcttaagt 190800
catgagtttg gtaagtaact acagaaatct gttctctcac agttctggaa actagaagtc 190860
tgcaagtcaat gtgttggtgg ggttggttcc ctctggagaa tctgttctgt gctcctgct 190920
taactttagg gcattgccgg caaccttga cttgtagatg catcacttca atctctgct 190980
gtccacaag gtgttggtcca ctgtgtctgt gtccagtgtc tttgtccagt gtctgcttt 191040
tcttctaata aggacattag ctatatgcac tcaccctaatt taagtatgat ctctattata 191100
cttgattgta tctgcaaaaca cccatttcc aaataaagtc acttttcccta ggttaggat 191160
ttgttttttt tttttttttg gtggaggagg aaaggaagga agtgggcaca gttcaattca 191220
taaggaacca actaaatata aatgccactg tggtaagatc agcttgacca tgtagatttt 191280
taaatgggtg ttttctttgt ttctattgog ttttttgctc atcttatcag aaaatttgat 191340
aacttttagt aataaatatg aaggagtaaa tcattaaaaa tttaaagtga aggccaggcg 191400
tggcggtcca cactgtatt cacagcactt tgggaggccg aggcggcg atcacctgag 191460
atcaggagtt tgagaccagc ctggccaaca cgtgaaact ctatctctac taaaaataca 191520
aaaattagcc ggtgtgtgtg acaggcacct gtaataccag ctactcgga ggtgaggca 191580
ggagaatcac ttgaacctg gagatggagg ttgcagtga cagagatcat gccactgtac 191640
ttcagcctgg acaacaaagc aagactctgt gtcaaaaaaa taaaaataaa aataaaaaa 191700
aaaaataaaa gtgaatccca cagatttgct tatattcttt ataactagat taagtgtcta 191760
caacagggct ctttttgact aacttttagg gcatatacaa caggtgccaa ccaactccaca 191820
ttccaaaaga aggcaaagac ttgctttgca ggacttgcaa agaggtgtcc aaaatgtttt 191880
cctggcctta ggctcagcat ccccccgaag gttatgaagg aagagctgtg gtctccctt 191940
ggagagattg gatggatgct gagtgtcat cttcttcccc ttccctgaca gacaacttga 192000
tgtgcatccc tggacttaca gggacaaaga agtgccgat ttattccaga aaatatttaa 192060
aggatgagag gcttgcaaaag aggccagag caaaatcaaa gaaggctata gcctgaaaat 192120
aattttatcc cccgatagc aggccagatt gcatctttct tatcactctc tgggtggacat 192180

```

tgtgagacag	agcttagcca	gggattatct	tagatcttgt	caaagaaggc	cctggccaat	192240
gaggcgatct	ctgcagaaag	cagctggaag	tgcatcatoc	cagatcccag	gttgagagac	192300
cagcctcagg	agaacagctg	agggtgaggg	attcattcta	cctgggaaaa	ctgcatgtga	192360
tggataacca	ttcgtgggca	gggatctcaa	agaaccccca	agtatgccct	gaagaattca	192420
gcaccaaac	tcctattact	ttcaaaggaa	tcaaacacac	gtttccaagg	gggcatgcca	192480
ataactttac	tagtgccgtg	ctctcccttt	ctctgtaaca	cctctcactt	cagtcctccg	192540
tgagccatag	tgtttcaagg	gtgtgtgatg	aagcgatgga	agcctcagg	ctggaaatag	192600
gaagaaagt	tactacaagc	ggcctgtgta	ttaaagtggg	cccaaacttg	tgaagaaaac	192660
ttggatcaaa	gttgcatct	tgattcttat	ggagtacaca	gaattgcaac	caataaaaga	192720
gacaattctt	gagaggaagg	gtgattagag	gacattttgt	tacctgggag	tgaacagata	192780
ctttatgaga	aatgctctat	atttgattaa	gtaacagggg	tttgctcttc	aaagtggggg	192840
tggttttaag	taggcagtct	gtggggcgcg	ggtaagaaaa	tggatttctg	attttatctt	192900
tgctcatttt	ccaggttacg	taaaatttaa	gaacattggt	aatttacttt	aataaagaag	192960
ttcataaata	gtccaatatt	aaaagttgat	taaataatac	ttttggaat	acatgaattt	193020
tactgaatgc	cctctagcct	tactacaaaa	agaatgcttt	gcaataata	aaagctaaat	193080
aaaaataagc	cagctccagc	cactttaagt	gacggaacac	actctgtaac	ttggggatgg	193140
ggacagggca	ggaatgaag	aaattaaagg	caaggcgcc	tcactgaggt	atatcaaatc	193200
ctattagacc	aggtcatttc	cttggtctgg	aactggacag	ggctcctatt	gtcttctcta	193260
ctttggctga	aatcttttat	ttatcctcca	catttggttc	atttgaggt	aattctattc	193320
ctgtttctac	atttaccatg	ctggatcagt	ttttccctg	attttatttt	taagtactta	193380
tgatattcgt	actaaatatg	taaatgttta	ttattgggtt	ttagctgttt	ctataataaa	193440
gagttgtagt	tattgttaagt	tgcatcgcac	taagctttta	aaatgtattt	ctccactgta	193500
gagaagtatg	ttccattgaa	atttgattgc	acacaaagct	tgacatgagg	tttattctct	193560
ttgtgaaata	aaccattcaa	tctggactgc	atccacattc	atctgaaact	tgaaggaaat	193620
aagtcattct	tcacttatgt	tttaatatgc	catatttcta	ttaaattaat	ggaagggttag	193680
acctaagct	caatgattgg	acacttatta	ttaatgattc	tttgacata	gagtgcatgg	193740
atagctcaca	agaaaatttc	accaaagtca	tcttatatat	ttctcataat	aatcctaaga	193800
ggtagggttg	atagacatta	gcattctgac	atcagaattg	gagacagatt	cttctgtctc	193860
atttaaggga	tttgcccca	aattaatcag	tccatgataa	agccagctct	ataatattac	193920
tctaaagaat	attcaaaagg	tcccgggcac	agtggctcac	gcctgggagg	ccgaggcagg	193980
tggatcatct	gaggtcagga	gttcaagacc	agcctagcca	aaatggtgaa	acctcgtctc	194040
tactaaaaat	acaaaaaaa	aaaaaaaaa	tagctaggag	ttgtggtggg	tgctgtaat	194100
tccagctact	agggaggctg	aggcaggaga	atcgcttgaa	cccgggaggc	ggaggttgca	194160
gcgagccaag	attgcaccac	tgcaactccag	cctggcgac	agagcgagac	tctgtatcaa	194220
aacaaaaaca	aacaaaaaga	atattcaaaa	agggtgcatg	ggcaaggaga	ggtcatttca	194280
catcctcttg	aggaacacag	gtgactccag	accacaattc	tcctagatgg	ccggggggccc	194340
ctgctccct	ccttttggtt	ctagggaact	gattcctgaa	gaggggtcat	gcgccaacat	194400
ctagtttaga	gtttttctat	tacaccacct	ggctctttat	attaatttgt	ttccctgatn	194460
ttttttttta	atctcctgga	tcacaatata	atcgtttgaa	tttttggtgt	tgtgtatata	194520
agggctccca	agaaatgtgt	aataaaaagt	gttggtggga	aaatcttgct	aacagaatcc	194580
agtagattaa	taactaagaa	aataaggcac	acacaacaaa	agggtgtgca	tggctgcaag	194640
ttcaagagtg	ggaactgatt	agagtggtag	aataaataca	tgtgcatgtg	ttctgaatcg	194700
attagacata	taaaaagata	tctagattct	ttgaaaaatt	aataagaac	tttttttcta	194760
agcataaata	tgaagtcaa	gtataaaact	accaaacaac	agagaagtga	taaagtgcta	194820
gcataaaaaa	ttagcatcat	gattattctg	tattgtttac	atagatctat	caagtcaaat	194880
gttggttcag	agccattctt	tattcagtta	tcaaaatggt	agagccattt	ggatgagttc	194940
agagacttct	tctttatgat	aaaaaagata	attgtaaatc	aaaaaagagt	ttaaaggcta	195000
tactaattaa	acagctaatt	ggaaattgga	agatttggtt	ctccaaactt	ctactatttt	195060
ttattttctt	ttctgtttgt	ggaagtgttg	gaaagcaaca	ttttattctg	tgattagcat	195120
caaagagttt	caggcatggg	ttataggact	ggaccatgca	tacttgaagt	gtacttcaaa	195180
ccataacatt	ctatgagatt	togtgagcaa	ctacctctta	ttgataaaaa	gaattgctgt	195240
ttaaacctta	ttttaaagta	cttacttttt	gctttctcaa	atataaggat	gatgtactat	195300
ggattagtaa	tatcattacc	tatggaaaca	aaagatatata	aaatttactg	ttcactcact	195360
tctccttaat	ttccaagaac	attaatcttg	aatttaagct	taaacttact	ttaaacatta	195420
tctttattat	cttattttct	atgaagaaaa	tattccaacc	tatctactag	tatgccattt	195480
taaaaacatt	aaaagccagt	ggataactgc	aaaagaaggg	cttttaattt	tatataaaat	195540
aataaaataa	tttattctat	tgataattcc	aaactctcaa	ataatatctt	gctgcagtgc	195600
taaggaaatg	gacaaaggaa	gaaaaaacag	cttttcatgc	ctttcctgaa	acaattgcag	195660
gaaatttttg	tcacgctaca	attattcttt	gatcatgaaa	aattctttta	aagacacgat	195720
agccacaacc	attttatctt	ctgacttact	aagcagctctg	agggtgaata	gagagtttcg	195780
attccagggtt	atgggtataa	ttactaactg	caaaaatcct	attgcctttc	attaagtagt	195840



taaaaatattt	aattgagact	acagtggtgc	tgacacacagc	gttggtacat	gcataattat	195900
ccctacacat	acaggaacac	tgagggtact	tttaacaata	aaacagatca	tcaaaaaatt	195960
acatttttaa	tttaagtttt	gactttaatc	ttctccgtga	ttaaatacct	atctctgcat	196020
ttttctgca	aatgaggaca	ttagctaaa	acaaatgaac	cataaagtga	cgggaaaaaa	196080
aactttgcac	tgtaatatata	aatggataaa	tctttaatat	gtattaacca	ttgcactatg	196140
tctttatttt	ctaacagctt	gtatgagaaa	gaagaactgg	tgatgcagaa	agtcaaatat	196200
cgccagctta	gcggtgcagc	atcctctcct	cccacataag	cagagatcgg	agcatgagat	196260
aaaatacctg	gatgcccttt	gctccagcct	cggtgctcca	caactttgct	tctgacagag	196320
tacttatctc	tctgaccact	gggaaactgc	taaaagagag	aaacttgaaa	aggattggac	196380
ttttctgtgc	aaatctcttt	gaaaattgcc	acctccaggt	tagcttttaa	gagaagccca	196440
atatattagg	aacctgaaaa	taaaggaaag	tttgacacatt	gaaacatcac	ttgctaatag	196500
ttctattatt	ggaaacggaa	atatggaatt	ctagaaaaag	aaatgtcatt	atggggttag	196560
gaattaacca	cttgctactt	atgttaattg	gacgtgggtt	aaaaaagctg	caggaagttc	196620
ctgaaattat	tgaacaaaa	ataaagtatt	ttcccctgga	atctaagtaa	tgtaagtga	196680
attctttaaa	cattctcttc	aactttatat	ttcaaaattg	cttcacagta	atttcttatg	196740
ttatccatca	cttgctaac	tggaaggcaa	gcatttttat	tggttttagt	actatttagc	196800
tttatgaaca	gcttctttgc	tttggtgcaa	ttacaaacat	ttatttccat	agggatggaa	196860
agattatgct	ttgctctctc	tgccaggtac	caattgtcaa	tagaaaatag	agccaactag	196920
agtaacttta	gctcttcaat	caacacatcc	agtattaaga	aaaaaaatc	attgatattt	196980
gagacgtaaa	ttcaaatctt	tattcaatac	gcaagaagtg	taattgtggg	agattaggtt	197040
gggaaataat	aaaaagctgt	ataattagag	tgctcaggatt	tacaccctcc	actggaagct	197100
actatatcag	gtataactcc	gcaaaaattgt	agatgggttg	caactaatcc	ttgctcacat	197160
tttratgatct	attactatgt	gccaggtgtt	ctgtctgaaa	cagcatggga	agaggaatca	197220
gagacagaat	tgccacctat	gtggggcctt	gctgggggtg	agaaattttg	tttctcacat	197280
ttttgaaatg	taatactttg	atctgaattt	tactcctgct	taaggacacc	aagatgaaag	197340
cggatggcac	cctgcccccc	acctgggtgt	agtttggtta	gaaatgactt	tagttctatg	197400
aaagtggatt	cttagttata	aacataaaga	actatcaccc	gctagtaaga	aaggatttta	197460
ttgaaaagat	gttgggagtt	cattttatct	cagggattct	gcagatgggc	agagagacta	197520
gaaaaccaag	agcacgaacc	aggcataaatt	caaacttcac	aacagtgagg	gtaccagaac	197580
agcttctagg	gaacaacgcc	aggcacgact	gcctggaaat	tgctgggtac	gggtgcttga	197640
ttgtcactat	ttcctggcat	tgccagctgc	agaggcagct	aactaacaga	gccaggggtg	197700
catccggggc	tctctttatg	ggggcatcca	cagtgaggag	tagtggtttt	tgatgcagga	197760
aattactgtg	ttgtatcttg	ggttttctca	atttatacaa	caaattgtata	catacgttca	197820
gatggaagta	agaagtacca	catctgagag	cggaagaggc	actgtaatca	cccatcccaa	197880
agaagatgat	ttcttaggaa	aagaagttag	tttgggaaat	atattagatg	tattctcctg	197940
tctcttccat	attcttggga	atttgagcca	tcacttatct	ctgatgtgcc	ttgcttgaga	198000
agtggtttct	atactatatt	tgatgatgtg	gatctgataa	cacaatatcc	aaatattttc	198060
cagctgtgtt	aactgagctc	ctagaggcag	atctcctgag	aattgtcagt	tgtgtttatt	198120
ttccagttaa	gtctactttc	agagagccat	tattgaaaga	gaaaagtttc	aacagaaaga	198180
tgattattgt	taataattatc	gtggtaagac	cacttaattg	gagatctatc	ctgttagcac	198240
atttttacac	accactgttt	gtgttggtta	ctataggtgc	aatgttgcac	agcagacttc	198300
aagagcttat	tcatcttcca	taataaaact	ctgtaccatt	actgttgaa	agcaacttgc	198360
catttctccc	tcttttcacc	tctgcaacag	aatgataaac	ctggagagca	ttatgctaaa	198420
taaaataaag	cagttacaga	aggacacata	ttgcatgatt	ccacttatat	gaagtatcta	198480
aaataataac	gcttattcaa	atacacgggtg	ggattttttac	tgggaaaaac	gctcgctctt	198540
tttaaaaaaa	aattgagaga	gatttgcttt	agttgatcaa	aattataaag	taccggaaac	198600
tccccaagat	ggcatattgt	tgaagccata	gtgtgtgcac	aggtgttgaa	taggcccttc	198660
agagataaga	agagtaaaag	gtggtataaa	aaagagaaaa	agtgcctagaa	tcagtcattt	198720
gcaacattaa	aagaaatgct	ttatcttggg	ggagatcctt	tggccaaact	atttcaactt	198780
ggaaagatgt	agtaataata	aaagatttca	atacaaggaa	ataaagaaat	tgcatccgtg	198840
ctacttctca	aaagagtaag	aatttaactg	ataaggttaa	ttattataaa	attaactttt	198900
accatctaca	ttaaaattag	gaattttctg	agaatttgta	ttctttttgt	cctttcaata	198960
aagaactcaa	tatgtttaac	tctattttta	acaggaatgg	ctgcctatgt	cttgacgttt	199020
gaatgcaaga	ccctcaaatg	ctcaggaatt	atcaagtaaa	gtttatgtta	agaaactagc	199080
attctgttga	atcactcaat	aatggttatt	atttaggata	tgcaaaagt	aatcacaaaa	199140
acttagactg	aaagatactg	cttacaatta	taggtgggaa	ataccataaa	atacaacctt	199200
tattttattt	ggaactcgtt	ttaatgcaga	ataatcctag	cgctctat	agtaacaaga	199260
agaaaccaat	ttaaatata	tgctgattta	aatagtataa	ataataaaaa	tgacaatttt	199320
ttgttaattt	tcagactcta	ctagactatt	tttgcctctt	ctattcttat	gaaatattca	199380
gggaagatat	tattactttc	actttacaaa	tgggataact	gagaactcaa	taaattgtat	199440
aatttatctc	aagacacaga	gctgggtcat	ggtagagcta	tcatacagaa	tctcacgatt	199500



attagttatc cctttccaat tctgaagtc tctgggaagt gtcagaattt acactcctca 199560  
 ctaggaaacta ctattgaggt tggaaactctg aggcctttatt tgggagtttt tccgtgattac 199620  
 agaaaatctc ttcttaagag aggttttaaat aaacaaaata agtcggggat attaccagat 199680  
 agccctata tccctcagagc tcccagcatc acacctttta taagaacagt atagaaaaac 199740  
 aaagtctcctt cccctcctct gccctgtgtt atttttccaa aaaacaaaca aacaaaaacc 199800  
 taactacaca ctaagtgtga aagtgtgttg agtgggggaa ggcaagctgg ggctgtacca 199860  
 ggtcaaggggt tgaaaagaat taggggatta ttacagttct agttcgaaga caaacacttt 199920  
 tcaatgtatg tcaatgaact ttattgaaca gccgccatat actagtaacc ctcttaggtg 199980  
 tgtgttacac ttacacatcg aaaccaggggt ttcccagatt tcttatgatg aaacagatag 200040  
 aaggtgataa tgcgtggcaca gcacacatgg acagtgtgac ggaaggagac caattaagga 200100  
 cgangcgggt ccgtatttct acacacctgc caccagctgg acggctgtgc tggaggtcta 200160  
 agccctggat gcttagctgt gtgatctgtg aaccggcagc atcagcgttc ccaggggtgct 200220  
 ggttagaaat acacctctca ggcccaccca gaccttctga atcagaatct gctgtttcca 200280  
 aaattcccaa gtgattctca tgcaccttaa agcttttagaa acactgttct aattagtggt 200340  
 cttaaaatat tgctatggga tacaggcttt agtaaatct caggatattc taccatgtac 200400  
 tcaagtttga gaaacaggag atttaataaa tactgatgtt tggatcccat ccatagagat 200460  
 tctgttttga ttggtgtagg gtgtaatctg gcattaacat ttttaaatg cccctgggtg 200520  
 gtttcatgaa tggcaatgtt tgggaacacc tgcttttaagt taagccatga gctaattgtt 200580  
 cttcatcata ttgcagatgt gggaaactgag acgttataag gaaatccaat ttgcgagtg 200640  
 tagagccaga cttaaagctta gacccttaac tcacttatct tcaaaaactgg tctgtgctgt 200700  
 ttattatttt acaggagagt tctatattgc tactgtttat acattagtat catttaaat 200760  
 tattttaata tatgccortg atggataata ctgagtgctc acttgattgg attgaagat 200820  
 gcaaagtatt gatgctgttg tgtctttgaa ggtgttgcca aaggagattt gagtcaagt 200880  
 gctggaatgg cagacacacc ctttatctgg gtgggcagca actaaatcag ctaccagctc 200940  
 ggccgggata taaagaaggc agaaaatcat gatggttaag caaactaaat atggcctgag 201000  
 aaggactctg tacttctata ggtgagctct tgtggatgaa ctgtaacct gcttaatagt 201060  
 cagacaaaat ttaaaacctt atttaatagt atgcactgta acaatggctg aatgttgccc 201120  
 agtcccagtg ggattcaacc actcatagac tgcttttgtt caaactgcgt tcaataaagg 201180  
 caaacgcga gctggaacca atctcactgt ttctgtacct cgttctaat tccgtgacct 201240  
 cactttacct tttttgtcta taaatttatt ctgaccacaa ggcacccctg gagcctctgt 201300  
 gagtctgctg tggttctggg gactgccgga ttccggaatc gttcatttct cagttaaact 201360  
 cctttaaat aattcagctg aagtttttct tttatcaatg aaaagactag actggcttag 201420  
 ccttccagcc tgcaccttcc tcccatgctg gatgcttctc gcccttgaat atcagactcc 201480  
 aagttcttca gttttgggac tcagactggc ttccctgtct ctcagcttgc agatgcccta 201540  
 ttgtggggtc ttgtgttctg gtgagttaat acttaatacc ctctccttta tatatacatc 201600  
 ctattagttc tgtccctcta gagaacctg actaatacaa tgtctccatt gaaaagaaaa 201660  
 gaaaatctat taataaactg atatagtcag agccctcaa gaaaaatacc aaagaaactc 201720  
 caatgttgta gataattatt ttattgtctc aagacatca aaataattac atgccttgtt 201780  
 tacgatgctt ttaagttgat cacattcaaa gatgtgtcct atccattcat gtacatacat 201840  
 tcaatttcat attgactctt gcctttttca gcagtgatca gtaaaaatta tctaaaaatc 201900  
 aatcctgaag gttatacttt atttttccat tttgaacctt gcccttgaaa gatcatttat 201960  
 atattttaa cacattttta aaaatgaatc tccagaagtc taatttctca tgctaaccac 202020  
 attagagagg ttaattcgat gacgaactct tagaacaat catcattaga acttgggtat 202080  
 tatttcaaac gcactctctc tccaaatctt agcagttgaa acaatttcac aggttaaacc 202140  
 agccagcttg acagccctg aatggctact cactaaaggg tccaagcaga ggcaaatga 202200  
 tggctctatc tottaaagac agagcaacag aatatttga attaaataga actggactta 202260  
 aatctcatcc cagtggacaa ctgggaaaac ttggaaacat taatcctgct gaattttgt 202320  
 gcactattta ataactaata catactgaat taagtattta gcaggattta caacactgtg 202380  
 tataaggagc ctggcatgtg tttcctgatc tttgggtgtt ttagagttaa tattatgaca 202440  
 gtgtgtaaac ttggtgtcga gaattaaaat ttgcttaaca gttctagtgt tctgtacact 202500  
 gttacctatg ttcactgcct ccaactcatt gaggctcatg atgctaaggt gaattctgga 202560  
 ttgaaatcta ggtctgtctg cttatagggc ctttttccaa cttcaggggt tccgtctttt 202620  
 gctcaattga tcccttcaag tctccatcac rtttgagcta cagcaaatca tgggtgtccrc 202680  
 tgttcaactga tgggctatgt gccagaagtt ccagaaaaag taaaggggga taaaagaga 202740  
 gagagagaga agaccataa atattctgaa gtatatgact aatctagtat aatggaagct 202800  
 cagaaaaata attgtatttc caaatatata tttctgaatt ttttggtatt tgttattgg 202860  
 attctctgaa gggccacttt cttttcttcc taatgctcta tgtccttaca aaggaatgac 202920  
 tattaattca ctggaataag aatcacacta attggcacac agtcccaatc tgcttggcat 202980  
 gtcattttaat gtgcagccat ttacctggtc aagtgaatac ttgaaaatta aacagtttcc 203040  
 ttgtgctatg attgatggca atgcagctca ggggttaagtc ttaaacgtaa tcccttatag 203100  
 gatttccctag gttctccctc agagtttcta actccataga tttgggggtg ggtcccagaa 203160

```

tttccatttc taacacgttc ccaggcgatg ctgaagcttc tgggccaggt accacacttt 203220
gaaaactgct gttttaatgc ttacatgggg tagcatgagt tagttttgcc tggggctttt 203280
aattccaagt tctgctactt aacagctgtg ttataatgat taagttatgt gatgtctctg 203340
tagcttgatt tctttgtata atgaggaaaa taatcatwcc cttttacata agattattgt 203400
gagaatttaa taaacaagaa atttcttata atgcctggta catagtaatc actcaataag 203460
ttatgtcaat tgtcaaagca catgtctgaa ttatagatga gagagaaata ttgtgaattc 203520
tatggcactt taatggcctg ttctctctac ccatctctag ccgtacctta cactggctta 203580
tcctgtgctc ttcctttcct gagggaaacc aaccacatat actgcctaaa gcactctttt 203640
ggcacttcta acatagggtg aaggaggaaa tttcaagaaa ttatgacaaa ctacaagact 203700
caaatctcct tttaaatttt aaaaatcaaa acatgtattc atatcccatg gcacatatgg 203760
atgaagatat tgtttttctc agttaatatt ttagttttgc ttttgtgcca gaggcttctt 203820
ttttttttgt aaatgtccct tatcaggcta ctgcagaatt gcacaaatac tgaattgccc 203880
caaaatattt gcctttaatt tgtttcttta aagcaaatgt aaaactacaa tgggtttatat 203940
ttacctagag gttttatcac catacgctta atttaaacct atataaaaaat agaattagtt 204000
atattttagt tttttttcat cattcactta atttaataata attcttagtc cataacttag 204060
atataacaaa aatttgtatt ttttctatcc tccatttgcc ctaaaatctcc ttaaacaaaa 204120
caaaacccaa ggaaacawtt ttgcctggaa aacgtgaagt ttgattagtt ccacctacct 204180
taatgcactt taatttgggc cccttctcct cggtagtcta gaataataac tttttttttt 204240
ttttttttta agagatctgt ttggaaaacc taccaaaaaa agacaaactc aatgattttc 204300
ttacagagag aagcctgggg aaagaattag caggaatttc tcaacgggtg ccgctgtctac 204360
tagtttttta tgcctgagtg cttgtcattt ggcagaaatc aattgttctc tttacaccac 204420
aagagtgggt gaaatggggg taaaactgca agttaagact gctctgaatg tggattcaac 204480
acttactgtg tgtcctccaa catgcgccct caaagtacca cttggctggg aattaaaaag 204540
ggttgccctt agcaaagctt taatttaaaa tagtcacagg tgacagctct attatttttg 204600
cacgytcaga tgcagccaca gctgttttta ttgatctcag ctgcatacac acccatttgt 204660
tgtaattatt agttttctgat gttagtgaag atacatatct cgaatgccta tcaaggctaa 204720
attaaattaa ggaaatcgta cttttcacct tcacctctt tggctctgct ttgctgcagg 204780
aaattgcac aatgtaaaac tttgagcctc atttcatctc cccccactag gagtgatttc 204840
caaagtgtgc ctaggttccct gatctattcc aatttaatat ttactgacat tttttccctg 204900
tatatttaac aactacaag caattaccat gtgrcaatta tgaatttaaa aattaaaaaa 204960
aacacaatta ttgctactac tgattaaaca ttgttaggac tttgaagaat aaaaagtga 205020
attcaacctg ccctcggtt tctaaaaatt gcctttttgc tcttaaaaaat gcatgtacag 205080
cttttttcat ataaaaataa tttcgagttt cttctatacg cataaaatat tttcagctat 205140
ttttgatttc aggaagtatg aaaaattaa gatacacaaa agaattgcgg catgaccctg 205200
tcgaamtctg tctcagtgac agacagtttt ggattgtaga cctcactttg ttttgcgat 205260
ttctgtttgt tgtttcacct gtgggtccatt cactgtattt gatgtgttg ctaaacattg 205320
cataacggyg tcccaccag taatgtaaga agaaacatgg aggaagcaag taagggtcta 205380
ggaaatttgg gagagaggaa aagagacata attaatgagg agagggatca agaagattgg 205440
caggattttt agaagggaat ttgatttgca aggtaccacg aacatggcac acaaaattgt 205500
aagaaaaaca aatttttctt gaacatttat ataaagattc ctttgggaaa atagaacaaa 205560
tttggttgta gagtggctg ggggttgatt taaattggaa aaacacagtt ttgtttttt 205620
ttaatttgaa ataagagctg atatgcaata ttgtagggga gttggttgac ttttttagtca 205680
gaatagtggc ttttacaacc atggtgacaa cattaactgt ggatcctctt gacgtcttca 205740
taagttcatt aatgctgggg gaaaatatgc atagcaacta atctaccaat atttgagaga 205800
attaggcaac ctctcaagaa tgtaggaaag aaatgtaagt caacaagctt atacaatgta 205860
aaaaccatga ataggcaggt aataaataat agaaattgat tttttctgtc cagtaaagac 205920
catgaaaaga gaaatattat aatttttaact tgagtgtctc agaatactgg gtttgtttt 205980
ttaattgatt aatatgattt taaaaaatta tcatgaattt aaaaatattt tctccattac 206040
tttgagttat cactttaaac atargtaata tcaagatatg tttctgaatt atatatttaa 206100
ttatgaccac taaagagcac ccatgaaatg gaagtgttaa cagcaatgtt ctgttgtagg 206160
aagataatca tattggttat gtctgcttta tagtaattgt attgttgatt gatttgcctt 206220
ttcctaactt aagtcctttt ggtcctttct tctattttgt ctaagagacc ctgttcaaag 206280
ggacatttta ttttcctttg tacaacagta tttgcctgac atttactctc ctctaataaa 206340
taatatataa ctctagtcct aatacattta cttacctaag gaaatcgccc ttgtgatgga 206400
ccccoggagg tgagatacat ttctggatga ggatctgttg agaagttgag catttggtca 206460
gctactaagt cactgccacc acctcgttag ctaaacatac tctgtgctca ttccttcac 206520
agtaaaattt ggttgggaca catgrcaggt gataataag gttattagct gcaatgctta 206580
ttaaaattt tgaagatata ctttagcaca ttttaataat tattgagact gtcccttagt 206640
aacagtagct cttctaaatg aaagctttta aaattgcatg taagtcaaaa caaagcatta 206700
tttgttattc aactttatat aaaagctgtc cagtgtattc aaatatgtac taacatttaa 206760
tttcaattca aaataagtat ttttatgttt ttataaagtg attaaatcac tttattgaat 206820

```

tctatccctt	tataaatgaa	taaaggaatg	aactcaagtt	gtttctattc	tattagcctt	206880
tacttttccc	ttcccccaat	taaccccaag	caaaaatgta	ttgactttga	cctgagaaga	206940
gatacgtcaa	ttttgttatt	tattttgcct	ctactcctat	ttcccataca	tatatgggac	207000
atataatca	gcacaaatta	gcaggcccac	caatttctta	ctcttgatc	cagtctgatc	207060
nnttctaact	cgtgaaataa	actaccatgc	agcacggatg	atcattcaaa	ttacagctta	207120
gggttgatca	taatgtcttc	attttagtat	cccccttagg	ttgacattat	ttattaaaaa	207180
ggtagttttc	aatcggattt	atggggcatt	atcgatagta	agtctacttg	atacattgca	207240
catacctttc	ttctgccaaa	aagaagtgtt	tacaaaaatg	tatttgacca	aatactttac	207300
ggtaaacatc	acrgtctgaa	agggaaactat	cttctttgaa	aggattaaag	cctccagata	207360
gatacaata	agattttttt	aaagtagaca	catgttaaata	catggaacag	caaaggggat	207420
actttatgga	aattgaattca	caagataata	atttgaattt	gacaagggtga	taattaaagt	207480
ccccagatag	atttcttcca	catctgtatg	cctttcagta	aagttgggtta	gcatattaac	207540
tactgctgga	gaatacagaa	aatattcatg	ggatgtaaaa	ctgcaaatta	ttattcattt	207600
caatttgtat	aataagtgtg	acaaaatgca	gtgcatacag	tgttacacaa	gaggtaccaa	207660
actccattat	agacattttt	agagcaaacy	tgggatccct	atatgacaat	ttttgtcttc	207720
aaatattcaa	gacatttaag	tatgaattac	cactttttgc	aattttaaag	aaatttttgt	207780
tatcttcaat	gctttctagg	gctaacaaat	tcttacttca	aatgactaga	atcattttta	207840
aatattatgt	actaggatat	gtggatatga	acatatatat	ttatacacia	atatatgtac	207900
acagccaaat	aaacaaataa	ttagtgtaat	ctatatgtat	taatattgnt	atattttcct	207960
ccagagaaaa	attgaggttt	gctgaccaga	gaagaagaaa	gcatcaatgt	gtgattaaaa	208020
aaaaaatgta	tatatatata	tatccactac	atttctaat	tgaatatgtg	atcctgcata	208080
ctaattctct	tattcataat	tgagtttcat	agctacttca	ttttaattta	gacattttata	208140
tgagaaagca	tgtttaatat	gatgataaga	aactgtatga	aattttggga	aagaactgtt	208200
aagtgtgccc	acacactgta	ggcactgaat	gtcacacttg	agaatgtccc	tggtcctagc	208260
ttatctataa	caggaatatt	ttccrtcata	tgggaatttaa	agataagaaa	tgaatcacca	208320
cgcagagaga	tgtgctatca	catcccactc	atggaaaaaa	caaataaaaa	aaaaaaaaac	208380
atgaagaaag	actttttgaa	gtgcagcatc	acctgcattt	atgatatttt	cagacatata	208440
aaacaaatat	tattctcatg	gttttgaata	aaattgtcac	ttttaaaatt	tgacatggga	208500
tatcacaatg	gacaatgcag	ggacccttca	gcccatagtc	aacagggtgat	caattgtgaa	208560
aactgatatt	ctctattagg	aagggttagct	agagttaggag	ataaaaaaca	acatgattac	208620
aaaaaggagg	atcattttgag	accaacttta	atgtgaaatt	gtgtagaaat	gggggaaaaag	208680
atacagaaaa	catcttcaca	agtccacttg	agaaactaag	ttggtgcaaa	agttattgca	208740
ctaattactt	ttgcaccaac	ctaattgtaa	gttatgcttg	aagaaactta	aaaggagaga	208800
cggagagagt	acacaactat	aaaagcaagt	ggaaacaagaa	gtgaaaatga	ccaaagaaca	208860
tatttgttga	acatggccat	tttgacaatc	aagggagaat	agtaatttaa	agtacaatgc	208920
cacaatctgg	agaagactgc	tcttgcacac	aagtgaagay	tctcatcggt	cctgtctgat	208980
gcaaattaca	ggtacattgc	tgacacaggtg	tcagatcgca	catagctggg	ctggcactgc	209040
ctctcctctg	gagaggaatg	cagcacattc	tccacaggcg	actggaaagt	ctggattctg	209100
tcacttagaa	atacacctta	gtwatccctt	tggcaaatga	atgggaaaac	aacctcttca	209160
cttctgtgtt	ttagaatttc	tgttttatgc	cttttcattg	tttcatgcaa	ttttattttt	209220
tcacgaggtc	aggcacagct	aactgttgat	tttcacccaa	ttaccttggt	attggcttgg	209280
ctccatggag	cagccccctt	caagaaagtt	tcccaagaaa	ctatgtactt	cttctacttg	209340
gactcagctg	tgtacctatt	tctctcgtca	gccattagag	tgttccacta	atatagtcaa	209400
agcagaacat	ctggcctatt	atataaatga	atattagtgt	tgaagaagat	atctttggag	209460
acagtagaac	aaaatggatt	ttttttccta	cacaggaact	cttggttttt	cctcactgcc	209520
cattgtcatt	ttttctacct	tactcagaac	atgattttcc	ctcaactgca	cctttctgta	209580
actttgacct	gtgccatcct	tgtgaaataa	gtgatatgtg	atttattttt	rtttctcctt	209640
ctaagtgaac	atgcagagtc	cttcagagac	catctgtgcc	tcctgctgag	tctctgtgtt	209700
ttctaccogg	gctagcttcc	tccagagccc	agtctcaggt	agatctgaac	cacacatgga	209760
ggagatttta	agaacaagtc	ttgtgacact	tcctatgctt	tgggtccaaa	aactattgat	209820
ttctctattt	ccttgcaagt	gtgtcagatt	ctggcaaaaac	tagcttggag	tttactgagt	209880
ttaatgatac	tgaatgacat	caggactcac	tcttctcttt	tcattttttt	gctatgagtg	209940
ggaaaaaata	taactgtata	agcaatgtac	tttttttctt	aacgcaattt	aatctttact	210000
aaaatttaatt	ttgataatca	tatatgtaca	aagtgtttta	ttttgttact	ttttaattct	210060
tagacattga	cttttaaaaa	tcttaatcgt	tgtctatttt	acccctgact	ttatgcccag	210120
tgtctatatg	cataaacaca	aatctttcta	ctcctatctg	tttagcaaaa	tgtaaataaa	210180
atgtctttta	tttattttcaa	ttttgtgtat	ttttttttaa	tttacttgcc	tgggttgagc	210240
atcttttagt	atatattttca	ttgtgtttta	ctgaaaattt	gacttaagac	tcagtgtatt	210300
actgtcaaac	tctccogtca	tgactatctc	aggtagctga	ggtatcactg	acagcataag	210360
sagcgacaaa	ccctcggtga	atgcaaatgt	gcatatccga	caaccccaag	ctaacaagt	210420
cataaccagg	gacttgcaaa	catgttttta	ggacataaac	tcrtcttaag	gttttactta	210480

aagtascatc	catctcatct	ttaacccagcc	tctcctaate	tttgcagaat	tgtccattct	210540
aatgcagctc	ttcttcattc	tttcacccgt	ggctccatta	aaygtttttc	tgcaggtagc	210600
ctctccattg	gagtaactgg	actgaggatg	gactcccttc	ttaagaaatc	cagaaaagca	210660
acaaaaatct	ttccagttac	tactggaaaa	tttctgcagt	catcatacat	gtatcattag	210720
gttctcacct	accattccta	gagtctatct	atctattttac	ctctgagggt	tcttagatta	210780
gccaaacact	tttgcagctc	tcgtattttgc	attgtagggc	cttggctgag	actagatctc	210840
ccaatatttg	gatataattga	ggactggaag	taaagaaasa	ctgccacttc	acacctatgg	210900
aaatactacc	tttttccata	gaggcagctt	gaatacatga	ttaggaattc	actgtgactt	210960
tggktttagg	gaaaaaaamc	aaaatgggtt	atttgggttat	atagagaagg	aaggaaggaa	211020
ggaaggaagg	aaggaaggaa	ggaaggaaga	aaggaaggaa	ggaaggaaga	gaaagatcct	211080
ttctttttcc	ccccaggaac	atggcagttt	gagaaaaatcg	ggaataaaac	awgtctttata	211140
atttgttaaa	tattgtaaaa	tactttcctt	attgttagta	gaattccatc	ttggcagact	211200
atttcttcta	ggaaaccatt	tacccttctc	cccttctctg	ggcactrcat	ttctctttct	211260
ggtaagaaat	ctgtcaaagg	cacttcacca	ggaatctggc	tcttataaaa	tgcartaaaa	211320
ctgcaggcaa	tggggaaatg	accattactg	aaactccttt	ggagasggct	tagcttatga	211380
agttttttgag	cagttctacc	tgatgaccag	cttactatgc	tcactctttac	ttcctcctgt	211440
aacaggtaga	ctgacacgtc	tcctctttgt	catgtgacct	tgcagagcct	tcaggaattt	211500
atgtcgactg	cacattttcc	ttggccagtg	gctctggctc	tggccattga	cttgcacatc	211560
tcaacagaaa	tgttgacagac	ctaacatgag	cagaggctct	aagtaaactg	gtgtgatttg	211620
gattggcctc	ctgtgcttct	gccattttgt	ggagcaacat	tattcatgcc	gccacaggtc	211680
cccaaatgag	aagatgcata	cagctgatct	agactttaagc	tatagcctag	aatcacatcc	211740
agcccatccc	aacagaccca	gagtgagccg	aaactctccc	agcagtttca	agagcagga	211800
atggctgttg	ttgcaagcct	ttgggtttcc	tggttatgta	taaagcagca	ttactacta	211860
aaaacttcat	aaatacaaa	ctttctgcat	gagtttctat	tctgtgccct	gcaagctttt	211920
attaaaggca	tgtagagtga	aagtctgatt	ttacctgaat	tttaggaggt	cagggccaag	211980
gctggatgca	gcctagcagt	tgtgctgtgc	catgggaagg	gacagcatgt	gccccaaaa	212040
aaagtccagaa	caaaaaaac	cacaaagt	gagtggtgaa	aagtgctaca	ggcttctgga	212100
atcagttggc	tcactgaagt	tcaagcagca	actatgaaaa	agttgtggga	catattagca	212160
ttttataggg	attcttagaa	atacctaggg	agagaaaccc	cataaaaggc	tagattttct	212220
gccataatac	acaaagggac	acaggctatt	ggtatttagg	cattaaaggg	aagtttgacc	212280
cacccctgaa	ggctgaataa	gtgtgagaca	cctggtttat	atttatataa	atttttgcac	212340
aggcaaaatt	caaccttcta	agcacaatgc	aaagtgcata	atccaatcta	ttaagcacca	212400
tgccctctaa	ctcataataa	atagcatata	gcaatcctgc	cacaggcgct	aacatttatt	212460
gagtttttct	actgactagg	ctgcgcctaaa	atatttttca	ggtttttaact	catttactgm	212520
ttcatgagat	aaattatttt	actatcttga	ttaaacaact	aagtaaagtt	ttgttttaga	212580
gagtttagta	gttacaccaa	aatccagcag	aagaaaatga	ctgtaatttc	ctatataaca	212640
atcctttgat	agtttttctt	tatcccttct	cottacatct	gtgactaaca	cattataact	212700
ccttaaaaata	aggtgggagc	cagcattttat	aaaatatatg	catatagaga	aaactaacac	212760
atacattaag	gagaggtgga	gaaaaaaaga	aagatcatat	ataatcattt	atttttatgt	212820
ycattttatgt	aatttatcta	tgtatataaa	tgacttacaa	taaaaatgag	tgtattccta	212880
gaaataatag	acaacagggg	attgtattct	tggggtgaag	tctacaatgg	atagatggaa	212940
tagttgtatg	aataacattt	ttgagtatat	tagcttatct	tcaaaatgtg	cttggctctg	213000
tcatttagat	tatatattat	ccatgtatat	aaagaataat	gatattaata	tactcttact	213060
gctaaattta	tacattacta	aaaataaaaa	gtgtcatata	tttgagcctg	aggacttatt	213120
tgaaagtctt	acatgaagaa	tattagaaac	atagggatta	cgtatcttga	ggatgaaatc	213180
tgtataattt	ctacttcagc	aaatcttgaa	aagcactctg	atagtctatc	tattgcctaa	213240
taatgctgat	catatggaaa	ctggatcaac	ttgcaaaactg	gatcaacttg	caaactaata	213300
ttacatgtag	taattattagt	ttacttcaat	tatctttcct	aactctcacc	ttgaaagtac	213360
tccactcatt	aaactatgca	tatacatata	tatatatata	tatatatata	tatatataga	213420
gagagagaga	gagagagaga	gagagagaga	gacaggatct	agctttgtca	cccaggctga	213480
agtacagtg	cacaaccttg	gctcaactgaa	acatccgct	cctgggttca	agcaattctt	213540
ctgcctcatc	ctccccagta	gctcagatta	caggcatgtg	acaccacacc	aggctaattt	213600
tttgtatttt	tctagagac	gggttttcac	catgttggcc	acgcttgtct	tgaactatgg	213660
aactcaagtg	aaccggcctc	ctcggcctcc	caaagtgctg	agattacagg	catgagccat	213720
ggtgcccggg	ctattcttgt	actttgggtg	tctttttttt	tttttttgag	tagtcataca	213780
gaattataaa	gtgattcaca	ccctcttatt	ctaaatgtgt	tcagtctaaa	tgyaaaagtt	213840
ttagtataaa	ttaattctgt	atatttttat	ttaattttta	gagtgagag	atattagcta	213900
ataaaaattag	ctttggaaat	aaaaagaaca	gatacggtta	tagttccttt	ccccaaatgt	213960
tttaaaacta	caaaagcaat	taagaagcag	tatcacaaga	tttaaaagttt	tttaagatat	214020
tattttgttc	cctaatagac	tacacaccgt	caatagttta	agttcactaa	aaatccagaa	214080
gttgccacat	taggctgaac	tcacatttat	tatttctgtt	attagcttaa	cttatattta	214140

ttataggagg	gatgacttct	ttttacaatc	atttaatatg	acaaataata	tattacattt	214200
tgatttaagt	acaactttgt	gctttatcta	aaaggtagca	ttatttttgt	gtgtgaatta	214260
atctatgttc	caataaaact	tccaacataa	gtaaggctct	gtatgtatgt	tatgttcttt	214320
ctatatctct	ggatttatatg	gatttacata	cttgatgcc	agtacatatg	tacatatgtt	214380
taagtataat	atacatgtct	gagtataata	tatacattct	tactgtacaa	actatgcata	214440
tgcataatcca	tattgcatgc	agctatgtag	atagactctg	acagcataga	aaaattttaa	214500
acttaagcaa	agttgaatag	ctctcaagta	ctgatatatg	tatgtgtcca	atattatgat	214560
tccaccattg	ctaattgcaca	caaaaaaac	taccaattta	tttttcttat	ataattatat	214620
ctctgataaa	attattgttt	attcagaaac	ctaatatoga	ccattaatgt	tccagtttca	214680
gaaaaacaaa	actcctaaga	gaatgaatag	aaattttctt	aaggaattat	aaagaaccat	214740
ctcagaattt	tactctgtta	ctgcaaaatc	tcatgcaaa	ataaccaagt	acggtagtgy	214800
aacgtctctt	cttctaaagc	tcatgaaact	gctctctgat	tcctctagcg	ttgacaraga	214860
atctggagtt	cagtactcca	tttcattcct	ggacatttag	cccattattc	ttgtatccat	214920
atttcatatt	ttcctttcat	gggggataat	tgtaagttaa	aggtgaattt	gttgactcaa	214980
accttataag	gaatcataga	tcaacctaaa	gatttctctaa	aaatatacct	acatattgaa	215040
tacttgcttt	gtctaagctg	cttaacaaca	acagaaaaaa	acaagcaaac	aaaaactcaa	215100
aaacaaaacc	aaaccaaagt	aaaacctgat	ggcccaaact	tgtaaagtgt	attttcttgg	215160
ttaattattt	aaataacaaa	tatgaggaag	aaaggcattc	tagtgaatct	gctttaaatt	215220
acataggaga	gagtccttatg	taaaactaatg	ataatgaata	tggtcggagg	gtttgtatgt	215280
ggagggtctt	gtgttaagca	ctgtgaatac	tttgttccat	ttaaaccgca	ccttagactc	215340
tacaaarctg	agagagatta	tgacagttga	ccaagattac	acaatgagaa	agtaagagat	215400
agaagtggaa	cttgatcaca	ggaagtcact	tctaaagact	caacatttga	ccctgggtctc	215460
taacgcattg	attttgagcc	cagcagcatg	tgaatttatt	cctatcagga	taaatgaaat	215520
gcaagtaact	tatttaactt	gagtcctcaga	tacatccaaa	taaaaataaa	tgccctcatt	215580
tatagcatca	cattctcagt	actaagtaag	ctatctataa	tgatacaatt	ccattgaggc	215640
agaagacttc	acttcatatc	tacacgacat	ttcattcttt	ctttctattt	gttttttcta	215700
gaagaattaa	ctgaaattca	tcccaagatg	aggcatcttg	gctgaggtga	gatatgcaaa	215760
tcatctgtgt	tgtttgacag	tcacgggtag	ataaatcata	tcactaacia	aagtcagttg	215820
ttaaaacttg	aagcacattt	aataaaaaaga	gagatttctcc	cagtcacaca	ggccaggtg	215880
gtgcatagaa	atggcatttt	atcaacttact	gaccagagag	ttgcaaatac	aatcccatca	215940
cctaaaaaca	aggaaacaaa	acaataataa	gaaaactttt	ttttttaatt	tttaatagaa	216000
agtcagtttt	gcttctgagt	aggatttcta	aaataaaact	atgggttctt	ttaacttttt	216060
ttgctttacc	ttactagaca	ctgagttatt	attttttgaa	ggaaagtga	ttataaatca	216120
gaaattattt	tctctttttg	ttataaacct	taggacataa	gaaatacata	ttccagagag	216180
aatttataga	aaaaataaaa	gtaatgcttc	tattattagt	attctcccta	aaactatagt	216240
tagaaattat	aaaagtaggg	aagggtccat	caatagtatt	gattgaaaat	gtaatgttac	216300
atgtatcaga	tacataaatc	cataaatcca	taacaatggc	aaagggagac	atcattacta	216360
tttgataaaa	ttaatatact	tgaatatatt	cccacataaa	ttcactaact	attcaaaaaa	216420
cttcacagga	cttttgaaat	tccgttggtt	gttttactga	ttatgacagc	tcacaaagca	216480
gtatgtgtaa	ctttcaataa	taatttagtg	atgaaatatt	accaaacaag	aggagaggt	216540
tttgattacc	tatgggtatta	gtttatatca	ttgtgtaaca	aatcttcata	aaaatataat	216600
tcataaaaaat	tatatttcat	aaaaatattc	ataaaaaat	aattttataa	aacatattac	216660
cacgtgttat	cagctgggtc	ccagacacag	agtcctgcag	agtcagtgga	agatgttggc	216720
cagactacat	tgtcatcaga	aggattggct	gggaagaatc	tgcttccaaa	ctcattcagc	216780
tttggtagaa	gttatttccct	tgtggttgca	tgactggtct	ttttgctggt	tgtagtgtga	216840
ggtccacatt	cgttctagag	gctgcttgcg	gtccttagag	gcctcatagg	caaggaactc	216900
atagtttctt	gccatgaaga	cttttccaac	atggctgctt	acttcttcaa	gcctgtaagg	216960
agatccrgtc	tcagttggcc	aggctctgtg	aaaggtaacc	tcatcaccgg	agtcaggtgt	217020
catcaccttt	cccatgttcc	attggttaga	agtcaccagct	cctgcccaca	ttcaagggtta	217080
aggggttata	caaatgtgtg	accaccaaga	agtcagagatt	atgtgggtca	cttaggtgtc	217140
aactgccaca	actacatatt	tgatataaat	ttcaaagaaa	gactgtgaaa	aagcataaag	217200
aattgttaga	attgcctgag	aattgaggac	gttcatctgc	aatgcctgtc	aggcttacat	217260
taaaacagtc	tagcttactg	agaaagtgtt	aaaaaatggg	tttgatagac	tttacttatt	217320
gggagaagcc	atttgacttg	tctttgaggg	gaaaaataat	ttaaaaatac	agatatactc	217380
tctctgcaaa	acaattcctg	gaatatgac	taacagaaaa	gccagcaga	cctgattcaa	217440
acctgatgtg	aacacagaat	gctttgaaat	gagtaaaatt	aattattaaa	cctaattttt	217500
caggtgagga	aactgaaaca	ctgggtttaa	gtatcttttg	cagggtcaca	cagctaaaga	217560
attttggagc	aggaattcaa	attcaggcag	ctggcttcag	acctcatgct	attaaacctg	217620
atattaaact	tgacatccat	cttcatgca	aatgttactc	tctctatcag	gaaaataacg	217680
catttataaa	agccagctca	gatgttactt	ccaaccaaca	aatttctgtg	acacactaag	217740
taatagagtt	aatcacaccg	tctgtgaagt	actaccocaa	gtatccgggc	tctaccacat	217800

tgtaatgcaa	taggtgtggt	cacactgttt	gtctctcctg	gtgaatcggt	aattcatcag	217860
agcatgactg	tggtgtgtgt	ccattatcc	ttagtaata	atgtagtgcc	ttgtgcatag	217920
gaagaaatta	atcttgatgg	ctttcaagtt	ccaccccaaa	cagtcaagcc	atgttgattg	217980
aaaagtatct	tctcagtttg	agagattacc	atcagagtat	ttgaaacaag	attgtatcct	218040
tggggactcc	atattgagga	cattttcttc	ataggattct	aaggctaat	gacttggcct	218100
gatcctaacc	taatccttta	gaaggctaga	gaataaaaatt	aaatatgttt	ttcacttttc	218160
cagtggagta	agcagggttt	tattgtgctt	cttgaatcat	tgtcacacta	gtaatctctt	218220
caaaagcgaa	tttgttctgg	aaaataatgg	gttagaggcc	tctgtggagg	aatacagtag	218280
tggaaatattt	tgaataaatt	gtaaagcaaa	acagataaatt	gcgtttttaa	gctctttgaa	218340
atatatacgt	aagatgttat	gtgcctgagg	ttgaacaaaa	agataaaatag	ttacagtcac	218400
taccocaata	tgtaccaatt	acaaatagca	ttgtgtgtac	atgtcctcta	ttataaaaatc	218460
agaaacgggt	gaaaaagatt	ataatcacgt	taataaatttg	atttgcttca	ttcatctttt	218520
gagttatctt	ttcataaaaat	tagatactgt	gttattttcct	tctggaaagt	acaatgggtgc	218580
atgtgaggag	aatgtttttca	tgttaatcaa	tgctgacaag	aatattacac	ttaaaaaaca	218640
cagaaggcag	gcagggttggg	ggttggagga	cagacaaatg	aagaagacca	gattgaaatt	218700
attagttgcc	agaaatata	ctattgactt	tttagatttt	tagaagtacc	tctgttagct	218760
tcccaaaaata	cagctgttta	cagtgtattga	taagttcccg	tgcttttaag	cttattcaca	218820
tccattagtt	ctgccttttt	cccctgttaa	aaatctgaag	ctgtcacatt	gaagtagcat	218880
gatcagctaa	tgttcctgcc	agattcaaaag	catgtcccca	ggggtggcaa	aagtattcaa	218940
aatctcaaaag	aaatggcttt	ttaaatgttg	cctgttgaca	ttggtcactc	catttgctgt	219000
atcagaaatg	ttgaaataaa	tatcaggcaa	ttatagaaac	aggactttga	tgatcaattt	219060
tctataaata	ttttatgttt	agcccaaaaa	aaagaaggca	tttggaaattc	actattttgt	219120
ataaaacttt	tattttaaaa	tagtatgtag	tatatatttca	aaattgggtta	attaattttc	219180
tagttttcac	ttcaataaat	gattattttac	tctattccct	ctagcattag	attattttct	219240
tgcaaacctg	tttttttttt	tctttttttc	tttttctgt	aatagtttta	gagtttctgc	219300
cttaacacat	ttaggtatat	ttttaggaa	caaagggaaa	aggtatcagg	gagaggaaa	219360
gatttctaaa	atgtaaattc	caaagacagg	gaatgttaaa	agtgtttgcc	agtcctgttt	219420
cttacagagc	acagaactca	gatgctctta	taaagataca	ggataaatca	catcatttcc	219480
tgctccatca	tcagaatatt	attatatgat	ttagatcact	tttttaaaaa	agaacatggm	219540
cttagtacag	aaacaacagca	aaagcctggg	gaaggagagg	agtgcacccat	gaggagtcac	219600
tggggagcag	aagccagctc	atttgactga	tttgggtcgt	gtgcaaaaata	attgctaaat	219660
aatgcatat	cgttgagact	ccgggtattt	tcaaaaaccag	ctggcaaaat	tgtgtttatc	219720
tctaccctct	gctggctttc	acgggttctc	tgttctctct	ccttttctct	cattctctct	219780
ttacccta	tcctgaccac	tgtaatccaa	taatctaagg	ttttaggatt	tggatgacta	219840
aggttaccca	tggaaattgtt	tggaaatgta	gacctgtaat	ggagagggga	gaaaatgaag	219900
aaatagaagc	tgtacgaaaa	tagtcaatac	tgaggccagg	taccgtggct	catgcctgta	219960
atctcagcac	tatgggatgc	cgaagtgggc	agatcactgg	aggtcaggag	tttgagacca	220020
gctggggcaa	catgggtgaaa	cctcatcttt	agtataaaat	acaaaaaatta	tccaggcttg	220080
gttgtggatg	cctgtaatcc	cagctactcg	ggaggctggg	gcaggagaat	cgcttgaacc	220140
tgggaggcag	aggttgcagt	gagccgagat	tgccgcatgt	cactccagcc	tgggcaacaa	220200
gagtgaact	ctgtctcaaa	agaaaaaaaa	aaaaagaaaag	aaagaaaata	gtcgatactg	220260
agacctggaa	actgagagcg	ttcatgaatg	ctgcatccaa	agaatatattt	gtactacgtc	220320
ttattcatga	tacaaatcaa	gcaataaagc	acactctgtt	gttcacatat	gacactatat	220380
aggtcaacgg	ggtgagagag	ccaccagaa	acagacacac	aattttggtta	tcagatattc	220440
tggagagggt	tctctgaggg	agtggcatct	gaagaatata	cagatgttaa	ccaaatgaag	220500
caggtgggtt	tggggaaaaa	agtattccaa	ggagaggggc	aggtgtgagg	acaactaaag	220560
agcaagggtg	attagaatac	tttgatgaag	tgaaaaaaac	actccaatga	gtgagcgctg	220620
atgaagaaaa	gggtggagag	cagattattg	aggtccattt	agatcatggt	ttagatcctg	220680
gacagcatcc	ttaaaataat	aaggagccac	taaaagtttg	aagaagggca	atgacatggt	220740
ccaatttgct	tttttacaag	atcactctga	cttctgcctg	gagtagcata	taaagggact	220800
aagcacgata	ccaggtggca	cataggagct	gaagtrgctc	ttgtcaaaat	aaggtgatgg	220860
ggtggggagt	ggatttggtt	cctaggctgc	tacaacagag	aactgtgaat	ggagtagctt	220920
aaaataaacag	aaatctattc	tgtcccagtt	tgggatggag	gtttcagcag	gacatgctct	220980
ctctaaagac	tttaggggaa	aatctgttca	atgtccttct	cttagcttcc	agtgttgctg	221040
gcaatccatg	gagctccttg	cttgttgga	tatcactcca	tcatacgggt	gttcttctctg	221100
tatgtcttca	ccttctcttc	tccgtgtatg	tctgcctctg	tacctcttct	cctcttccaa	221160
caaggacccc	agtcataatg	gaatagagcc	actctaataga	cctcatttta	actgattaca	221220
tctgcaata	tctattttct	aaataaggtc	atattcatag	gtattgcaga	ttaagactgc	221280
aatatattct	tcttggggat	atagtttaac	cccaaacagg	ggattattga	aagggaagtga	221340
gataagggtt	agtggaaaga	aaccatgtaa	aggttttaaa	atatattttt	ggtggtagac	221400
tcaacaggac	ttcatgatta	agtctgagtt	tagggatgga	agcagagtcg	tggttgactt	221460

ccagggtttct	tccagatgat	gatgccattt	gttgagacac	aaaatccaac	ggatcaagta	221520
agtgggtgggt	tgagcatggg	ttcatttttt	tgacatgcc	agtttgatct	gtctgtgaaa	221580
tttttatgtg	gagctgtaga	tttggtgggt	aaaaatgcct	ttggaattca	aatgagagat	221640
ttgagttgga	gagataagtg	tgggagtcac	ttctattcaa	attatcatca	aagctatgca	221700
agtagataac	tttctccaac	cagataatgt	ttagtcagaa	gagaagaaac	cttgacatca	221760
acattaaaca	ttttaagaa	caatgtggac	tatttcagac	taatatgaac	aaaaaggacc	221820
agactaacc	tgccataaga	aaaaacaaac	cgtcaagaaa	cagacaggac	atttgaaacg	221880
atgggttttag	agacaaaaaa	catcaggcag	ccaaggacag	aaaaataaag	aaatacattg	221940
atcccaacaa	ttgcctgaga	ttacctaagt	tcaggtaaaa	gacaatatg	tctgctttca	222000
atattttctat	ttaatgtgat	attgaaatgt	gcattcatag	tagtaagaca	gcaataaata	222060
aaaggcatcc	aaattgtata	ggaggaaata	aactgggttt	atctgtagac	aatatcataa	222120
actatataaa	taactatgaa	acatacaaaa	aggtataaga	gctaattgat	gatttttagt	222180
aggtttcaaaa	attcaaatgt	aatatataaac	attaaaatca	cttgatacaa	gagcaaaaga	222240
caatatgtaa	ttgaatttta	aaaaattgta	cttataatag	caaaaaatat	aaaaatattc	222300
cagataaatc	tgacaaaatt	tgtgaaagtt	tcacacagtt	aaaaatagaa	aacactaagg	222360
aaaaattaat	gattacctaa	ataaatgcag	aaagatacat	ttttaatggg	ttggaagata	222420
atattcttaa	gatgtcagtt	gtccccaat	tcacttatag	cttccctgtt	atcccagcaa	222480
gctacattgt	ggaaattgac	aagctgaatt	taaaattcct	atgtaaacgc	aaaagaattg	222540
ttacagctgt	aacacttttc	taaaagaaca	aaactgggaa	taagtatat	tttaagacag	222600
ttattagcat	taagatagac	aaatcaatag	gataaaatag	aaagtaaaaa	aaaaatagac	222660
ctatacatat	atggacaact	gattttggcc	aaagggcaaa	ggaacttgt	tgaggagaata	222720
gcaatatcat	attttcaaca	aatgatgttg	gaacatctgg	gtattcatta	caaaacgggt	222780
aatctttgat	ccatatttaa	aatttaactc	aagttgaatt	atagatgttg	gtgtaaaatc	222840
tttaactata	aaattcatgg	aagaaaacat	atgagactaa	cagtgtaaat	tagcgttaag	222900
taaagatgta	ttagatgcaa	cagaacaaaa	catgatccat	gcaagaataa	atgtgataaa	222960
ctggacttca	taaaaattaa	acaatcgcaa	caacaaaaaa	taacaaagaa	atgaaaaaga	223020
gcttctgtct	tttaaaaaatc	attattaaga	aaattaaaag	acaattctcc	atggtacaca	223080
agaaaaagatt	tacaaaagggt	atatctaata	taagatttgt	atccagaaag	tataaagaaa	223140
cctcaatggt	caataagaga	aaaaaaataa	aaatgagcaa	aagctttagc	agaccactga	223200
aaaaaaatgg	gtagagaata	tgcatatgar	aagatgctca	atatcattag	tcattaggaa	223260
aatgcaaatg	aaagcataac	aagatagcac	tgccctcttt	tttaaaagtc	taaaactaaa	223320
aatgctgacc	atatcaagtg	ttggtgagaa	tatggagaaa	ctagaacact	catatcctgt	223380
ctgtaggaat	gcaaaatggg	taaaattactt	tgaaaaacaa	taatgaagca	gattccttaa	223440
ataaagggtg	atctatccaa	ctactattag	atctattcaa	ttgatctagg	tatttaccce	223500
tggaagttaa	agcatatctt	cattataaga	cttctacaca	aattatcaca	tctttactta	223560
cagcagctga	aacctggaaa	caactctaata	gcccgtcaac	agaggaatgg	atggataaag	223620
aaactgtgat	gcagtggaa	acgactcaac	gaagatgaga	ctaaaaataa	ttatactgag	223680
taaaagaaac	caaacaaaat	agagcaaaac	ctgtgccatc	ctgtttatac	cttactccag	223740
taaatgcaaa	ctaatacaca	atgaaaaaaa	ttacttattt	gagaactggg	gagaggaagg	223800
agaggggaaag	gggtagataa	agaaaagagg	agagattaaa	aggagcataa	gaaaacctca	223860
gagaataata	ggtttgtggt	aaacattacc	gtggttaatt	ttttagggtg	tattcacatg	223920
taaaaactta	tccaattata	catttttaaat	atgtacagtt	tagtgtgtca	gttatgcctc	223980
tgtaaagtgt	attttaaaaa	aagttcctat	ttccaagttc	acaatttcat	ttgaactctt	224040
aaaaactggc	caggagcagt	ggctcatgcc	tgtaatccca	gcactttgag	gagcagaagc	224100
aggcagatca	cgaggttagg	agtttgacac	cagcctgaac	aacatggtga	aacctgtct	224160
ctactaaaaa	tacaaaaagt	agccagacgt	gggtgtgggc	acctgttaatt	ccwgttccag	224220
gaggctgagg	caggagaatt	gcttgtaact	gggagggcga	ggttgcaagt	agccgtgatc	224280
gcaccactgc	actccagctt	gggtgacaga	gtgagactcc	atctcaaaga	aaaaaaaaaa	224340
aaaaaaactta	caccaaggtc	aaagcagagg	agctggcaaa	agagactgag	aaaaagagca	224400
gagtgataga	gaataagaaa	tgagagctcc	cacggaagct	aagggcagga	accacacagt	224460
ccttccctcat	gctacttaga	agttaaaaaa	aaaaatgaga	gttaaaaaag	gccattcggt	224520
tcagaagtaa	cattatggac	taatgaaaag	gaatttagtt	aaataaggca	tgagacaaaa	224580
gaaagttagt	tgactgagaa	atgatgaaat	agagtgagca	agtaatatgg	tttggctgtg	224640
tcctcaccca	tatctcatct	tgaattgtag	tttcacacaa	cccacgtatc	atgggagggg	224700
ccctgggagg	taactgaatc	atggggggcag	gtctttctct	tgctgttctt	gtgatagtga	224760
ataagtctca	tgagatctaa	tggttttata	arggggaatt	tttctacaca	ggctctctct	224820
cttgactgct	gccatataag	tccctttgct	cttctctcat	cttctgccat	aattgttaagg	224880
ccttctcagc	catgtggaaa	tgtgagtcaa	ttaaacctct	ttcctttgtc	tcagggtatgt	224940
ctttattagc	agtgtgagaa	cagactaata	cagcaagtat	aaaaaacatt	aaatgtttgt	225000
cttctaaggg	aagagaggag	agtagagaga	gagagattgt	tagctgaagg	gtaaaagtatt	225060
cttggttatta	agatatcaga	gcctagaaca	tatttaataa	ctgatgagaa	ggatacaagg	225120



gagagagaaa	atttgaatca	atagtttgtt	tccactacca	cagaaaagga	tgagatccag	225180
aggttttgata	gaggaatgag	ttttccctcat	tggagagata	agagcagatg	ggttttgaaatg	225240
ctggatccta	aggggtgggat	attaagggag	tctgtttttt	atggccttta	ttttctcttt	225300
gatcttgcca	gcaaggatgt	ctgccaagat	ttaaggggtt	ggacaggagg	atcatagttt	225360
taatgaataa	aaacatatat	gaaatagcca	cttcagagag	tggggcagac	agtatgtgtt	225420
ttagttttaga	ccagagtccc	tggagttata	ggggtttgaa	ttcaaatcct	agctcctcca	225480
ttcacctgtg	gtactgtgaa	caatgacaga	accaagttaa	gaccccaaat	cttctttaaa	225540
gtgggatgat	aatggcatgt	tcacccataga	cttcttggga	gtaataaatg	aaatacttta	225600
gttcaactat	tttcacaggt	cctgacactg	ggttttgaat	aaattgtgcc	caggactatt	225660
gctattataa	agcaaaatga	aaaaacacaa	tggcccaatt	aatatgtttc	caaggagtcc	225720
aatactaaga	agaaaagtat	tgttggaat	ggaaatccac	attactagat	taacatgtgg	225780
agaacttgct	caaagaggag	gctggtttag	atgctgtgat	aatttttagat	gtcaaatgga	225840
ctggattatg	gaaataccca	gagaactggt	aaagcatcat	ttttgggtgtt	tctgcgagga	225900
tgtttccaga	agactttggc	atctgctca	gtggactgag	tgaggaagat	ccctccttac	225960
cgtgggcagg	tgccgtctca	tcagctgatg	gccagatgg	aataagatag	aggaaggaat	226020
atgtgtttct	ctccttctta	gggctggaac	ttactcttca	ttccctgctc	ttggacaaca	226080
gaactccact	ttttggactc	taggtattac	actagcacc	cacttcccat	agccgagttc	226140
ttactcctta	gatcttggtc	tgagagttac	accattggct	tccagggcca	cactactcca	226200
tgccaggatc	tccagcttgc	agatggcttg	tctggaggct	tttcagtctc	tataatcacg	226260
agccaattct	cctaacaat	gctatctatc	tatctatcta	tctatctatc	tatctatcta	226320
tctatctatg	tctatctcca	tccatccatc	catctatgta	tctattatct	atatctatcc	226380
tattggttct	atctgtctgg	agaaccctaa	tacagatgcc	agcaaaccaa	aaccaagtgg	226440
caaagtgtcta	ccattgacat	tcctctcctt	ccttaatgac	tcttagtaac	caaaaacaca	226500
tgtaaaattc	catgtaatac	acaagccaga	aaaggtaaaa	tattgtatct	gctagttcac	226560
acagctcatt	tcactaccac	aatctaccat	tattcaata	ctgattttaga	aatgacagc	226620
ctccaaaaga	ctaaaatcat	cagttttttac	ttagaagaaa	aagaaatgaa	attttaatga	226680
cagggaattt	taatcttact	atttcaattt	ataagatgtg	actggctgaa	tattgttatg	226740
tactaaaatg	ataaatacta	tgtgttatta	aatattttgt	cattgaaaaa	tgacctatta	226800
ttctgacctta	ttttattttc	aggggtactt	ttagataaag	gagtttataa	aggcttctgc	226860
taaagaagac	atcagagaaa	atgtcttcca	attgatttat	cttgctttgt	ttctgagcat	226920
cttgactcca	actgtattgc	aatgttttta	aatgttttta	aatggcagag	actctgattt	226980
gtcctaggcc	ccttcttcta	aattccaaga	gtctatgtga	actcttctta	tgagtttctt	227040
aaaattcacg	aaagtagtct	ctactggttt	aaatgggtgc	ccctcaaaat	ttatgtcatt	227100
tccagaaccc	cagaatgtgc	ccttatttgg	aaatagggtc	attgtggatg	tacttagtta	227160
agatgaggta	atactagagt	agagtgggct	ctacatccaa	tattgctggg	gtctttctag	227220
gaagaggaga	agagccacag	ggaatgggtg	ccatgtgaag	acagaggtag	agattgactg	227280
atgtgcctgt	cagccaagga	aggtggcgca	ttgcccggca	ccatcaggaa	gtcggagaga	227340
agcctggcac	agattctctg	tccagtggga	accagcctgt	gtgttaattt	cagacttcta	227400
ccctccagaa	ctatgagaca	ataaattgcc	taaaactggt	ttaagccact	cagtttgcaa	227460
tagtgtatta	tgtcagccct	aaaaagctaa	tacaaagccc	tagagaattt	caataaaggt	227520
gtcaggaaaat	atagatcaca	gacaataatg	ataggacctt	ctagcattct	tacacttctg	227580
taatgtttgt	tgtctttctg	gtctctccta	tatgttttta	ttgtgattgg	caaaaaatccc	227640
ttgaaaaaac	caagaaacaa	tacgaacaga	aacaaagcta	tgtacttgtt	ttgggaagca	227700
acctcaaatg	gaaagatggg	ggctgtgttc	taaaaggtga	ggagaaattt	atagcgaaga	227760
taaagagata	aagataaatg	gtcctgtgca	ccactgaatt	ctgagaactc	tgaacaaaac	227820
ctaagtatat	gttagtgccc	agtgaatagt	tgttgaaatga	atggatgaat	caacaaataa	227880
atgaattgat	gagataatca	atctgtgaga	caaaaagcaa	cacaaatgag	tagacaagag	227940
tgggttagtt	tggctaaaaa	ttgaagatca	cccatTTTTT	cagccaagta	ggcaggcatc	228000
tgctacttgc	caggcatggg	tattgggtgtt	tggcataatg	aatagcaaaa	acaaagtcct	228060
actgtcaagg	aacatagact	cttgggaattt	agaagaaggg	gcatgggaca	aatcagagtc	228120
tccgccatta	aaaaaaaaat	atttttaatt	gcaatacagt	tggagtcaag	atgctcagaa	228180
ataaagcaag	ataaatcaat	tggttttaaaa	tatctttcac	actacccttc	catggaaggt	228240
gcagagaagg	ttccagtcta	aatgtaggca	actggagact	cccttttctc	atattccttt	228300
aaaccttatg	ccccaaaatg	tgtaatcaag	tctattcttg	aacagaaaaag	aacacagtgc	228360
cctttctctt	tggaaatttta	tataagcccc	taagaaaagc	tgctattgtc	ctagtttgta	228420
aaagtacaaa	tgggtgagta	tgcattctgct	ctgttcagtt	ggaaaatttag	tacagggtttc	228480
agcactcctc	gataccaatt	gactattgaa	tgcgactaaa	gggaagcttc	tgtacagacc	228540
agtaaaaccac	acaattgcag	aaatcctcta	attaaagaaa	cacaaccataa	tcttctgtgt	228600
gaatctgaag	aaaaggtttt	ttagaaaagt	acatctgtat	aaggtaacatt	tgtgtaaaaa	228660
acaatcattt	ttcattgaag	tgagctgtga	ttgttagatg	ttgttctaag	aaaactcagc	228720
ttaaaatttc	ttataggctg	ctttggctta	aaataattgc	ctaactggat	ggtcacaaat	228780



gaccaagaaa	ccaaatatat	ttcatagtat	taactatcat	tacagagtgt	taaataatac	228840
tgcacagtag	agatcttttg	aaaaatat	taaatgatag	attctatatt	ggtttgaggc	228900
aaagtcttaa	acataacaag	ttaaataaat	ttttaatgaa	attttacaag	tgaaataatt	228960
aatgttcaag	acatagaatc	atgccttttg	gaaatgactc	tcatgatatt	ctattatatt	229020
aaaaactatg	tgccctattt	taggaattaa	attgctctgg	tagatacccg	ataaatatct	229080
atctcacaaa	tcattgtttt	ctgtcatatg	taaatttttt	aaatgtttta	gtttataggt	229140
tggttggtc	tgccaccag	gctggacagc	agtgtctgtg	tctcggctca	ctgaatactc	229200
tcctttctgg	gttcaaggga	tcctccctcc	tcagcttccc	aagaagctgg	gactacaggc	229260
atgcaccaac	atatctggct	aatttttaaa	ttctttgtag	agatggcatc	taggtttccc	229320
tggttggtct	tgaactcctg	ggctcaagca	gttcccttac	tttggccttc	caaagtattg	229380
gattacaagc	atgagccact	atgcctgggt	agtttcta	ttttacattc	agtttatttt	229440
tatgtgactg	gcatttgtgt	aggattat	ttataaccaa	tatgttttag	gtatgtaacc	229500
cagaagtaaa	tcactaatta	atctgcaccg	gaggaagtat	aatttaggga	gagtagaacc	229560
tctatctgat	agactgcact	gcaggtaagg	tgccacagcc	tggaagcgcc	tcagtgaact	229620
tggaaaaaac	aagcaccag	actgaccttc	cagatccaag	gtctggtagc	aggatggttt	229680
tcatttccat	ctgttggagg	tcctcagctc	tgacataaag	gcctgcttta	tacaaaatga	229740
gcctgattga	aagatgtaag	ccaccagggt	agaggagag	ctacaggagg	cagacaggag	229800
aaagtaagag	aagagtgcag	ttgttcttcc	actagcttgt	ggctactaac	aaatttcctc	229860
agaaagtgtc	tgaggattcc	tcagaacaat	ttacagagat	aactgtaagc	tgagaggggc	229920
tatgaaaacg	ctcatgggtg	ctgacaggag	cccctgggtg	gtggtgggtt	gggaatacag	229980
gggctctgta	gggggtgaag	gaactgtgct	tgacaccaat	caatgggagc	ctggagaaac	230040
tgaggcagg	actggaatcg	tgctgagact	caaacgtttt	gagttagagt	actcaccagg	230100
tactctaact	caatcatatg	ccttgggtat	aaggaaatta	ggaataattg	gaactgtggc	230160
tgcccccaaa	tttcatgccc	tatttatagc	cattcatcaa	atttagaatg	aggtaaagga	230220
aggattcctt	ttattatatt	ttaaaatacc	cacctcattt	tttacacca	catcataaat	230280
caccgagagg	gagaatttag	cctgtggacc	tccttttttg	tagtacctca	tatgtaattt	230340
tttttttttt	ttttacagat	tcattgtttt	tggtacaagt	gaacgtttta	taacraaaag	230400
gcttttttgt	tggtgttttt	atttctata	aaaatacctc	tggttgaaaa	ttgttaagaa	230460
agctaattat	tgtatcttta	catgactcta	aatttgggtc	atttaaaaag	tcaaaacata	230520
gacatttggg	aatgactttc	tcctttcagt	actgaagtgt	agaggagag	ggaaaaaaag	230580
tgagaaattc	cagtgcctgg	atgttcccag	gctaatttct	agacaagtga	gttttgataa	230640
gttcaagttt	cagatgagaa	tttctttttt	cttctctcat	tggttttccc	acatcgtgag	230700
ttaaaatgtc	tctgaaagag	tctgaaaaaa	tatatataat	aatgtggaga	attattttaca	230760
taaaatttta	ataaactaat	atgagaatga	gacctccaa	taaatattgt	ttagagttaa	230820
tttaatagtc	ctctccccta	tcataaccca	tttaaaaatg	tggttatata	tatatgtgtg	230880
tgcatccta	aaaatttagt	gtgcagttct	tagtttttct	ttatgcaaaa	cacacaaaat	230940
aaatgtgaat	aatgaagcc	attacacaga	catttttcct	gctaagttct	gtaaatagat	231000
attccaataa	agacactacc	aacatcttac	tttgtcttaa	atataattga	gagtatttcc	231060
atattaaacg	ttttaaataa	aatgattgg	atttcaaatt	gccttaaaga	ttcataatgt	231120
aatattcttc	accctaaatc	accagatta	tttctaattg	tttctttatt	ttgccttcct	231180
aaatacagcc	tttgattaat	ttagcggaat	ttatcaatct	gagatatatt	tacottccca	231240
cattacagca	tgccatggaa	tgaaaactta	ctttgttgc	caaagtatc	aaacacagtt	231300
ctgtggtgca	agttcctctc	cttttctaaa	tttgtctaga	ggatctcata	aaacgttaact	231360
cctctgacaa	gggaaccatt	ttagcaccac	cactgcaaaa	gcttctgtgt	tcctaaggga	231420
aagatccttt	cctgaattaa	atttaacctc	tttagtactc	ccatttagcc	acctgataaa	231480
tcacttggag	ctatcttttg	ggaagagaga	ggtatctggg	aacaataaca	cttctttttt	231540
gaacagttta	ataaagcttt	gtgagatttc	aagatgaaag	ataatgtgta	atgctgatag	231600
tgccctccaa	ggctctgcat	tcattggatcc	aattacggtt	tttgtcatgg	taaaagccac	231660
agtggatata	ttaaatraga	gtgtgggtta	agaatgaagg	cccaggagtc	tgagatctg	231720
gtttctaagg	ctgacttcac	ttctgctttc	ttatctcact	aggggtatgc	agcttaacat	231780
ctttacctca	gtgtctcctt	ccataaagtg	cccagcaaat	atctgagaaa	ttcttttaaa	231840
atgacagtgc	aaattgccat	ttagaatcac	cttggaaatg	ctagaaaaaa	atgaggtgaa	231900
tgttcggtta	gaacgggaaa	tacaatcttc	tgcttaattt	cagatttatc	atcaagtttg	231960
tccttttact	tttaaaagtc	atgagatttc	cagaagagga	atgacatttt	caaagcctta	232020
aggacacttg	tgaaaacact	ccttcaacta	ttctttat	ctcaccacaa	ttttgactct	232080
ggcttttctc	caacttcctc	tcctccccaa	ttcccaattc	caacttgaaa	agagtaattg	232140
caatcttaac	tttttccatt	ctggagtatt	ttcttgcagc	cgatttggtt	catataattg	232200
taaaatttat	catgtcagat	ttgttttag	atgcacgttt	gctggcttta	gtcagtcatt	232260
ccccaaaccc	ttaaactgat	taatttcacc	gtgaaataat	tgattacgaa	ttgttagcat	232320
tatctgactt	tattttaaaa	tgcaagttga	tagaacattg	tcatttattg	attaatctat	232380
tgccctttgc	cttgagataa	ttttgaagaa	gcagaatagt	gccattagta	agacaatatc	232440

tatgtattac	aaatgatcat	gttcttaaaa	tacacacgca	cacacacacg	cgcattgcaca	232500
cactcacact	ttctagtaat	aagcagtggt	ttgagtaaaa	tcttttgaaa	gatgttttagc	232560
aaatcgaaac	aatttctgtc	ttcctatgtg	atgcttgcaa	ttatatgggt	tagatatatta	232620
ttgaaggact	tgggaaacct	catagtccca	ttggcattat	tctcaacata	tcaaagatttc	232680
attcagttata	aataaataat	ttcattactt	ataataataa	aagttatcat	ggaaaattttt	232740
ataacctagt	aaataggatg	tcatagaaat	attcaacacg	gataaataaa	atatatacat	232800
agtttaaaatt	agaacacctg	ttatagaatg	gtacagtcta	actctctggc	caggtaagct	232860
tctagcatga	aacttcacac	aatcatacct	acaatatgtt	cttgcctaata	tccttgttcc	232920
aaaggaaagt	ttgataagga	agttgaataa	tttagcttcc	ttttagagat	ggtagagctt	232980
atttaagtag	ctttcctgta	gtacaaaaag	atttaggcagc	tctgatttga	actctagatt	233040
gaaaatgagt	agaaataacc	tcttatagaa	cacgtctaaa	tttaactttg	cttcacaccg	233100
ataaactata	gcctctacta	ctttatagac	agattatagc	tcatcagtag	gactaaaaat	233160
cactgtggaa	ttactctgca	atagtcctgt	tgctatgtct	ataaatataa	tttcagttag	233220
tgaacttttcc	attttctagc	tgcttggtta	tacttttttt	gcgtggcaca	ctaaagggct	233280
ttaggaatat	ggtaaatgac	aaatttaaat	agcctcagat	aaaacatgaa	agtcaacatt	233340
aaattatttg	gttaaacatt	attttcagaa	caatttttaa	aaataatatt	ttgcaactga	233400
ttttaactta	atgtttatca	tattctaaaa	aggtaaaaaa	aatgcaatgt	tgtttattct	233460
cgttaagcct	aacatgaaaa	taaaagcatt	atttagggga	aaaagtccca	tttaattatg	233520
tgaacacata	attaaataat	tatcataaga	ggcaaaaaaa	gatagatagc	atcagatatt	233580
gtagcttggt	agagcttcag	tactaggggg	tcaaacagtt	gaaaaataatc	tgggaaacag	233640
ttttaagtggt	tttatattta	aaacacctac	acacacaaac	acacacacac	acacacacac	233700
actctcacgt	aggcagtggt	aaatttacat	tgcttcttct	ccatgttatt	gttgtgtatg	233760
ggcaggctta	tttttcacat	atttccctct	caatagcctt	ggggttctaa	agagagacac	233820
aacattatca	cccaagatgc	ttgacaaaga	tgggtatttt	gaggggtgatc	cagattttact	233880
gatgatatga	ggcctggaaa	tctgaaaatt	taacaaaact	tcctagtaat	ttttataaaa	233940
aacaaaaattc	aagaaaaact	actgcatgaa	aaaaatgtga	gaagccagtg	gatactttta	234000
aaaaagtgtt	aaaatgggtc	ctatgtaata	gacaaataat	ctctgagcta	gataaaaagt	234060
caacatgaag	tataatctaa	tattaaaaag	cccagagaag	ttgctttcgt	aagttccaac	234120
attagcattt	agctttctcc	tgggttaaaag	tttaaaatga	agaagataca	tactctgcac	234180
agggtctagg	aaggtaagag	acccaaggac	ataatgatct	ctgtaaatgc	ctctgtgcca	234240
tttccaaaata	gcgctctca	tgcttctctc	tccatccttc	tttcttgggt	accgtctctc	234300
ttccttcttc	tgtctccttt	gctgcaacag	tattttaagt	gatataataat	acataattaac	234360
tctatcacat	cttttaacat	acatgttggt	ttaatgactg	tggagcgatt	gtaggaattg	234420
tccatttttac	ttaatttttc	agaggttttt	gggttaagaca	aaattccctc	agagactcaa	234480
atgtcatctg	ttgattgaaa	tatacatctg	tggttttgaga	ttaagcccaa	acgaaggggtg	234540
gcttctctct	ctggtacgga	ctatactttg	agtcattttg	ttacagttaa	acaaaaatat	234600
caaaaatggca	cgaaaaaatt	aactgtggtg	cctaattctt	tgcttctcat	gaaatggatt	234660
gcctatacca	tacttaataa	acatgtgcat	agcctttcca	aagctccaaa	cctcacgatg	234720
aaatagtcct	ctcgtccctt	cctgtgtgga	gtgttgggaa	gattcccttt	ccctgtcaag	234780
gggcctttcc	tcoggtgttg	accacaaggt	agttttttga	ggggatctta	agttgtatta	234840
gtattttatg	agttctttaga	taggggttaa	taattctctc	tactttggcg	tggaaagaga	234900
atggtgtgtg	atagtgaatg	tgggaaaaag	ttgtgaaaga	ctgtgaaatg	agctgtaaaag	234960
attactaaat	cacctgacta	gcagagtctt	ccaagatttc	tctgcctttc	atcatctttg	235020
agctattttg	caactttgta	aaatttttaa	aatatgatat	aaatgcaatt	gactttcttc	235080
cagaaaaatg	caaagagagt	gtccagtga	aatgctaagg	attattgctc	tgtgaatcga	235140
ccagttttac	ctgtagatga	ttttaaattt	gcctttgaa	tggttgtaac	atcttattgt	235200
ttctcttatt	aatcaagtaa	aatctttgat	tcatgttcat	gatacatctg	tatgagctta	235260
ccatgtcttt	aaatgcctta	tcatgttaac	ataacctact	ttctagacaa	atagsactcc	235320
ctttcttatt	ttcaagtaag	tagcaaacac	ttatttgaaa	ttctgagccc	atcaatttat	235380
tttttatatg	aaaggatgaa	catttcctag	aagccctctt	ggatcatctt	aggtaaacct	235440
catatcaaat	ctgtgcta	atgttatttg	tcctggaact	ggttcttagt	ccaaatttta	235500
taccattttat	tcgactctta	tatccttcac	cttgatcaat	ataaagtga	tttgccacac	235560
attcggtttg	accaatatta	ttctctgtat	agtggtttta	actcaattat	caaactgtat	235620
tttaaaaaatt	gaatggccag	gtgtggtggc	tcacatcttc	tataatccca	gcactttggg	235680
aggccaaggg	gggaggatca	cctgaggtca	ggagttggag	accagcctgg	ccaacatggg	235740
gaaaccctgt	ctctactaag	aatacaaaaa	aattggccgg	gcacggtggc	tcatgcctgt	235800
aatcccagca	ttttgggagg	ccgagggcgg	cggatcacaa	ggtcaggaga	tcaagaccat	235860
catggctaac	acgatgaaac	cccactgtga	ctaaaaacaa	aaaattagcc	aggcttgggt	235920
gtgggcgcct	gtagtcccag	ctactcggga	ggttgaggca	ggagaatggc	atgaaccggg	235980
gaggcggagc	ttgcagttag	ctgagatcgt	gtcactgcac	tcgagcctgg	gccacagagc	236040
gtgccatcaa	aaaagaaaaa	agaaaaaaga	aaagaaaaca	aattagctag	ttgtggtggt	236100

gtgtgcctgt	agtctcagct	gcttgggagg	ctgaggcagg	agaattgctt	gaacccagga	236160
cgagaggtt	gcagtgcagc	cagatggcac	cactgcactc	cagcctgggc	aacggagcaa	236220
gactctgtct	caaaataatt	taaaataaac	aaataaataa	aaatagaata	ttttggttgg	236280
gtgtacatgg	gtgagctgct	ttattttatg	tatgtagaat	ttctaaaata	attgtgcttt	236340
taccgtctag	cttctataat	ttataaatat	gtaataaata	gtagataatg	ttaatgagtg	236400
tgttgttttg	cattttgtta	gttggcttaa	aaatgagaat	gataaaaaag	gcatgggtatc	236460
agggcatttt	aaaattatta	gccatggggg	taaaaaagaa	aaaggaaagt	gagacaagtt	236520
gccagcatgg	ctacacagag	aaattccagt	ttccccaaaa	ctggaaaaga	atagcatttt	236580
atatgaggac	agtgtgctta	atacataaca	aagatagctg	ctatatatta	gaaattatgt	236640
ttgttattcc	ttaagcagtt	atttttaatc	ttgatagttt	attctggtta	gatataatac	236700
aagaatataa	ctctagtctc	ttcttgcaaa	aaaaaaaaaa	aattccaatg	agcaagatga	236760
ttaatccagt	aaagaaacat	gtccagcctg	gggacaaaaa	aaaagtagat	caaatttttg	236820
ttaggaagat	taaccaagta	aaaatccatt	agctgggaatt	tattgtaaaa	gccaataaga	236880
aatcatccaa	gaaataaaaa	actaaaaata	gccctcaata	tggataaaaa	aaagacatag	236940
agtgattctc	tttgcatcca	atagagactg	agaaaacaca	actgtcagaa	tattaaatat	237000
tgctttggaa	agttttggaa	tatgatgaaa	gtaaattaga	caattgctag	aaaaatatcc	237060
ttcagtttgt	tccccacaca	tcacaaaatt	cagcttgata	ccattttctt	tttcttgaca	237120
aatttagggc	ttttattagt	gctatgacag	aaagaaggaa	tgctaccaaa	aagaccattt	237180
gaatttactt	taattatttc	agagtaggga	aacgataatt	ggaaagtagt	ttataacctt	237240
gagtttgtaa	aatctgtaaa	ctctcacaa	gatgtgaagg	cagagaaagt	atcaaaactc	237300
gtccctcata	ataagctcat	tacttcataa	aatatatata	cccttcacat	atgtgtcaga	237360
aatgtttaaa	attttcacca	ctgggtacatt	atttttgggc	cttacatttt	tgtggatttg	237420
gtgaatacct	gaaagaaata	aaataggatt	ggataatgtg	aatttctttt	acactcattt	237480
tcattattag	aaaataaagc	ccaaataaaa	tatgacagag	cagaattagt	tttaatacag	237540
gaaaaatttg	caagttgaag	aggagaaaat	gcagccagga	actaacggta	cacacctagg	237600
tggttgaagc	agtgtggatg	gtgataagag	aacataagga	aaattctgcc	ctccagctac	237660
tttgaattta	gtgttctgcc	tagtatgaat	attattgaac	cctatggtat	ggtacttaac	237720
tatctaatag	gatgataact	ctctcctacc	tgtagaaaat	gtagatgcct	tgtctacttc	237780
tgtttatata	catgctgaaa	catactgatg	tctgctgtct	tcttaatttt	acactaattt	237840
ttaaagtaaa	gattgatttg	ataaaaaata	tgcacttgga	aaaattattc	tgctattccc	237900
acatttcatg	ttcatgtcta	tttacagcta	acaagactag	cacgcaataa	gatataatgt	237960
ccaagatctc	gcttcttttc	atgtctggct	tctccctgtt	taccacaaat	tattttcata	238020
atattttact	gtgattccac	aaaccagaga	caaattctct	aagggaatag	tgttccaagg	238080
catggccatc	attagaagac	aataagcttg	ttttccttcc	aacattctta	gctttgttga	238140
aggctaagtg	gaagaaatta	acatactgca	tgcctttcaa	tgaaaattgt	ttctaattat	238200
tttgcctctc	aaattattcc	accaggcatt	gattagactt	tttaaatcca	tttcataatc	238260
gtcactctca	gctaacattt	gtaatttttt	aattgttgtt	aaagagaatc	tattaattga	238320
aatataatta	tctcttttaa	taaaagttaa	taaaattact	tgggtagcca	acataattcaa	238380
gtaatcatca	atcttgggtc	gaaattaaaa	gtaagatgag	catctaaaact	tggcactcaa	238440
agactgacca	tctgcacctt	tgaggcaaaa	taaaattcag	tttaggaaaa	ccttatgaaa	238500
atcaatgttt	taattatgct	atcaagggct	gtggtaggca	gaaacttctt	tgtctcacc	238560
cccaaaaatg	tccagttctt	aatcctgaaa	acctgtggaa	tcttgtgtaa	caaagggact	238620
tcatagatgt	aattaatggt	aacgaccggg	aggtggagag	tgacctcgtg	tatcctgggtg	238680
ggtccaaggt	agtcacaggc	atctttataa	tcagtgaagg	tctccctgct	gtgcttagag	238740
agggggatgc	tggagaaga	gtggtcagag	aagcaacatt	gctgactcya	aagacagagg	238800
gcaggccaag	agccaaggca	tgcaggaggc	ctccagaagc	tggggcaggc	aagagaatgg	238860
actcagccct	ggagccctca	gaaggcatgc	agccctgccc	attccttgct	tttagccagg	238920
taagatgtct	gtcagcctga	tgagaactat	gaaataatga	atgtgcttgt	ttttttcacc	238980
actaaaatttg	tgtttatttg	ttatagcagt	gatagaaaac	taagacaagg	ctatatcttg	239040
gcatgtcaga	ttttaagtta	tctgctatag	tttagaaaa	cccanagaag	gagctgcaat	239100
ttgtctcttt	gaaaaaaaaa	aaaaaaaaaa	ggatcacctt	ctaggggatt	aaacattaat	239160
gggtctgggtg	ttgtggctga	cacctgtaat	cctagcactt	ggggaggccg	aggcgtgct	239220
gcactccagc	ctgggcaaca	tagtgagacc	ctgtctccag	aaaaaaaaaa	aaaaaagaaa	239280
aagaattaaa	cgttgacaga	ttattaaagt	aggcactgtg	tgagacctag	cttgggtttg	239340
cgctataga	tatttaaaag	ttgatccagg	ttttcacttg	aaagcagaac	tactgcctta	239400
aattcccaact	tcacaaaaat	aatcatgga	ttgagggaaa	aatttacttg	tttgctttat	239460
aaaaagagat	atcgatctct	tcataagtta	tatatgtttt	acaaggttgc	actgtttgaa	239520
ctgtaagtgt	gaatggttgc	taacactcca	gaaccaaga	ggtacatttt	aactggaaaa	239580
attttcttcc	tgtatcttac	aaattaattg	aagtttgtct	acgcttgaaa	atatgtcaaa	239640
agtattatgt	agctttacca	cagttttatc	taacttctag	aagaatttaa	attcatataa	239700
agataaagag	taacatacca	atagagttaa	gagagttttc	agatttgctc	agaagaatgc	239760

tcytttgaaa	acctatgaag	tagattttac	ccaatgtata	tctggatcta	aaaattaaag	239820
aaaatatcag	tcgtgttcag	attcctttat	accctgcaga	agagagaaag	aataagcaac	239880
atcaatggag	catataacat	gtatggttta	gcactttcat	atttttaatc	ttaactattc	239940
agccaaatac	gtcttatttt	tttattttta	tagatgaaga	tattgaggtt	tgcatttttt	240000
ttttataaat	tgctcagtoa	tacagtgggt	aaatagctga	caggagagtt	cactccaggt	240060
cttactggct	ctcaaacctt	tgctttttta	aaatctccaa	actagcgcc	ctccctttta	240120
actgcaaaat	aagcacaaag	ttatcccagt	tctacacagt	gctaaacaaa	ttcctgtgag	240180
caaacaatta	gcacaggcag	atggagtgag	ggggaggtgg	tttataacaa	tggaagagctg	240240
tctgctgtgt	gtggattaaa	gattcaaaact	ctttgatcct	cctcccaggg	ataggtgag	240300
gcctatgtct	cctcctcatt	aacatgagtg	atccttggt	aggttctcct	ttgactgaca	240360
gaatatggga	gaaagttaagg	ctgaaccata	tactgggctc	aggcttccaa	cctgacatg	240420
ttccacttcc	tgtctccagg	aacattccacc	ctgggagctc	tgtgctgcca	tgcaaatgtc	240480
acattgcaact	gaggccgcca	tgctgtaaaa	aaccccaagc	tacctacctg	aaaagctgga	240540
gagaatcagc	acccatcagc	ttttccagca	ttttgggtcc	tcccagctaa	atcatcgcca	240600
ttgtgaagca	gagatgagac	ctcctgccat	gccccttcca	aatcctggac	caacagaatc	240660
atcatataca	ctagttaagct	tttactttta	gccactaatt	tggggtgtag	ctcttttgca	240720
gcaaaagatc	atcagaacac	tgccatgcag	gagctgataa	acttgctcac	tgttatgtat	240780
gtatccaatg	actactcaca	cacacattac	ttgcttcaact	ctactttctg	gctttccatt	240840
cagcaaat	taagaaatga	aaatgttatt	cgtgggaaat	taactcaaga	tcaatattgt	240900
tgtagttaat	ctctttgtaa	atggtaaaaat	attaaattat	tgtgagtcac	ttttatttaa	240960
agaagtgttg	ttctttaaat	gaaagaagtt	tgggtctttt	gtggatactc	tgcatataga	241020
cattctagac	atgaaagaat	gttgaagaca	aagaatgttg	aagaaatata	gacatgaaag	241080
actgttgaa	tttggttttt	tgtggatact	ctgcatatag	acattctaga	ctggagtgat	241140
acatctgtgt	caggtttcta	tgtttttaaa	atgataactt	gacacgccta	ccaatgccaa	241200
tttttaggat	atatatctgt	atatctgtgt	gaactaagga	aaattcagga	atgattaaac	241260
ctacatat	tgccaggaga	gactggataa	tgtaagtctt	tatatatttc	acaacttcaa	241320
gtttatagca	tccaaatgag	caattgctga	catacctagg	ggatttctct	gcaggttcaa	241380
gtctgatctg	gcctgagggc	aacactctta	ctgagtgtgt	tcctaggcac	tccttggcag	241440
ttggctccag	gccacaataa	agagccctga	tgattctctg	agcaggccca	acttccattc	241500
acattcatgg	gaatatgagg	atgacttgtg	atcttcagga	cttagataga	ctctttcagt	241560
tggtgtagg	taagctcagc	tcactcttag	agacttccca	cagtggagct	catgcatgac	241620
tctgtgcagg	tgccacagga	cagccagaga	caggcaccta	aggagtgcac	taccagcagg	241680
tggaactctc	actgagggag	gtagacaagg	gagaaaagcca	ggtgtccaca	tggaagcagg	241740
cttactcagg	gagactgaga	ttttgctcta	cacatcataa	aacgtggagg	tcattaatgg	241800
aatttttcac	aggacatgac	tgccatagcc	atgatagggt	ccctggctaa	tctttcaatt	241860
rtgatagcat	cataagataa	ttttgaatat	gaggagagaa	gaataactct	agaactgcaa	241920
tatccaatat	ggtaaccctt	agcaaccctaa	ggctacttaa	actaattaaa	attaaaactt	241980
tagttctca	gtgccactag	ccatatttca	agtactcaat	aaccacacac	acttggctac	242040
catgttgga	agcactgaca	tagaccattt	ccattactgc	agaaagtctt	cctatttgac	242100
agtcacctc	tggaaccaat	ccggagtctt	tagggacctc	aaagactgag	gtgctaaatc	242160
tcagagcagt	actctgctta	atgcttagtt	aaaagacttg	ctaacttata	tagttgatct	242220
agagaacttt	aaaggcagct	agttcgtagc	aactatgttg	ctggaccttg	aaactgtatg	242280
tcaggaactt	ggtgttctag	tggtgacttc	tcttttatct	accatgttca	cacttttctt	242340
accactcatt	tatccattca	gaaagtattt	cagaagtccc	tatgatttac	caggtgctat	242400
ttcagacact	gaggcactgt	ggggaacaaa	acaaacaaag	tctcctttct	caggtatctc	242460
acattccatg	tggtgatgca	gactgtaaaa	attcacacac	agaagatact	ttcatacagt	242520
aggaagtgtt	atgaagaata	tttcaaaagg	ggtgtggggt	gggagaggtt	gatgttcagg	242580
tggttgcttg	agaatgaact	gttcatacaa	agaatcaata	attcatttgt	tccacaataa	242640
tttattgagt	acctgttatt	agcaggctac	tttattaggt	tgccagagca	gggttttctt	242700
agcacaatta	aaacacaata	acatttgtta	gaaattcaaa	gacccatgtc	cactcttgca	242760
taatgaatca	gagtctcgag	gttatggctc	agagactggt	gtatttataa	atcactgaag	242820
taatttttaag	gtatagccag	aatagagagg	acaccggcat	tcttcttgaa	actcaacaca	242880
ctagacaatg	agaaatcatt	aaaatgaagt	aaatatattg	attgagattg	ttttggcttc	242940
aaaattaata	atagcttgga	gaggaaagaa	ggtaaattat	aaaagacaat	taagaatctc	243000
ttaaatgtat	ttagataata	gttacagagg	gacaaaaatg	agtaggatct	gtgataacag	243060
aaagaataaa	aataaacaac	aggacttaaa	aaagtaagaa	taagaaagaa	gtcaacagta	243120
accagaaaat	tttcaacctc	atgtctaaga	gaatgatgat	gctatttagc	aatataacca	243180
aaccaaaaat	actacgagaa	gtagggtgtg	ttcactgtgt	atgttgtaaa	tttataatta	243240
ctttttcttt	attttggatg	catgcttttc	ctttttctga	agttctgaga	catacaaaaa	243300
caataaagaa	aaacatgatt	ttatgccaga	aaataagata	ggaacaattc	ctcatgctca	243360
atataat	agcctccac	ataatccaaa	atatatgtgc	agcaaatgtg	gggagtttcg	243420

aaattagatc	caacttctta	ttacagttaa	agataaaaag	aaaacaaggc	tttgtctttt	243480
ttaggatcct	ttttggaact	tgacagtataa	aaataactctt	tcaaaaaaca	taactcaaca	243540
tcaaaagtcca	ataaaacagt	aaactggact	aaatatattc	gtatattagc	ttgacctag	243600
tctgggaatt	gagccagact	tttcaaatta	tactatcatt	atttacagaa	aaaggaaatg	243660
ttactaaca	ttattaaata	aaagcacata	tacacatgcr	taacaggaaa	caaaatgaat	243720
gttaatgtta	acaaatgttg	aagatgtgtg	tgtgtgtgtg	tgtgtatgtg	tgtgtatata	243780
aagcagataa	gatattctct	tgacatgat	tggattggaa	atagtagcta	tgccagggga	243840
aattttatgc	aatcccttg	gctgtgggct	acacatttat	tgtgttatac	aaatatgcca	243900
aaaaaaatgc	catttacaca	gttgccataa	acaaaattga	aattatacat	tagtgtttaa	243960
agtttttttt	tttctttcgc	ctccttttct	taatgaatgg	cagataatga	aatttaggag	244020
tttgcaaaaa	caaacgtagt	ttaaattgta	ccttagttgt	ctatgaaata	caaatacatt	244080
ttatatattt	ttggaacat	ttgtgtgtgt	tcagaaagtc	catattttgc	attcaagatt	244140
gacctgtttt	taagtcatat	ttcattttgt	acactcaaat	tgactaaagg	cagatacttt	244200
ggcattcact	ttttgattca	ttattaaggc	attattttga	ttgattctgc	ttaaggtagt	244260
tttcttgatc	agagtagacg	tttaggccac	ctatctagac	aaatatccca	ctttcctatc	244320
tgaatatatt	gatctctcct	tacccattct	agggttttgt	gcactctctaa	gatgatcatc	244380
ttgtgtgtgg	gtcctgatct	tccagtgac	ttagcagcac	agtaatttgg	ctgaaatgtt	244440
ggtgagttca	gacaagaaga	aggtgatgat	gacaagagag	cgtcacgggg	aaattctgcc	244500
ctgaggaagg	aatagtttga	ggctggagag	atgtgctacc	cccgcactc	tttctcaaat	244560
ggtatttatt	cttgaagac	agtgtaaagt	gtctgaaatt	gcagaaagca	agtgttgct	244620
gtagaataac	ataccatgac	gtgcttcaac	tcgtttcctt	aaaggttaaa	gaaagaaaaa	244680
tgtgcactct	gaagaaaatt	ccaaaccctc	attgtgagga	gtaatccaat	atttcaattt	244740
agaaattcag	agggttttaa	ataagaagca	attattctag	gcaaaaagca	aatataaatg	244800
ccagcagtca	aaaaaccagt	gaaaagatat	gattaaaact	gagatgtaaa	aaatagatat	244860
ttatctgaag	ttcttgactt	ttaaagtcca	caggcatttt	cccttaagaa	aagaataaaa	244920
acattaaatg	ttaaagaaaa	acataagaaa	tattataaaa	aagataatta	ttcagaaaaa	244980
attagtattt	tgttaccac	accaaagaca	aacagagtat	ttttcaatga	ctggaagcca	245040
gtttaagtga	ggaaaaaaaa	aaacacgtta	ttattactgc	tattatccat	tatttttatt	245100
tttacttttg	gcagtgttat	ttgctaatac	gctaaccgat	ggtagtgga	ttgatctttt	245160
atgtattata	aaggttgtat	cttctcattt	taagggaata	ttttaactac	attgtccaat	245220
tgtacactat	aaatgcagca	agaccaggtc	aagtgaact	tcagtgggag	agtctctttc	245280
ttttcttgtg	tgtagggtct	tcctctccac	ctgctacctt	gactttttag	attccctgtt	245340
ccctgggtct	gtgcttgtat	aatatacaag	ctgctcttga	catagcacgt	tcctcttttt	245400
tccttcatta	tttttaaatg	gttataatat	actcttaaaa	acaaatttga	atattactaa	245460
tagattctaa	tattatctta	ctctctgaaa	ttaagttagt	atgtatattg	tgttaaacgc	245520
ttgtctatgc	tgatgcaaca	tatgggcagg	tctctctctg	ggcctatatg	tatttgtttc	245580
atatcggtgt	acacagatto	attcattttc	ttaaatgggt	atataatatt	tcagtgtata	245640
aattaatcac	ctgtaagtgg	ctmgtaat	aatttgttgc	tacttctctc	caagatgaac	245700
agtgctttag	tgagcatcta	gccatttctg	tattctgtcta	cttattgatt	tatgcatata	245760
ttcaccacct	tcttcttcca	caaataagac	acacatcctc	ttcatgagag	tctacacaaa	245820
aatgttacaa	ataacatttt	taacgcaata	gttcctttaa	gtcacactt	gtaatgtttc	245880
tgccagact	taagccagcc	aaatgtgtgg	aaataaatc	catttgtgta	ttaaatatca	245940
aagataggga	acattcataa	caaatrtaaa	satagccaag	gattcattga	aacattttcca	246000
cataccatct	ctccaaatct	ggttttaaag	agattgtaac	ataatttcag	gaataaagac	246060
cttctggaat	tgagatttat	ggaaaagaga	ttatagctcc	acaagattct	agggatgcaa	246120
aaggtctgga	ggagcctcca	ggcacaggca	ctttgtttcc	atgctgaaga	ttcaatgtgg	246180
gttctgaaaa	tgctgcttta	gcaagcacac	atttgttcca	ctggtgttgg	attttctatt	246240
gattaatagt	taccattttt	ttcttctctt	aggaaagaca	acctgtaggc	acagaaatca	246300
ttgtctgtct	tacactgaaa	tctcctgctg	actttgagag	ctgagcattg	gatttttgatg	246360
ggagagtga	acttagtgaa	gaaatgattt	gaacctttgt	tgatctaata	cttctgtgat	246420
acagtatatct	aaacagaatg	agtctaaagt	cagtataaag	acctgtgtct	ctccatgttt	246480
tgctgcttat	aattgcagag	gaacattttg	agatagaaaa	ttttatccct	ctttcaagtg	246540
actttttata	aaacataacc	ttttaagaga	ggtcattctt	ttgaaaaatt	gtctgaatac	246600
aaaaaagtta	ctgaacagca	agggattgta	gccataatgt	tcttgaccac	tttagattga	246660
caaatagtat	ctcaaaattg	atcaggttct	ggagaatcta	ttgaggttgt	atggagtaac	246720
caagtgatgg	aagaggttgg	aagtgaataa	ttttatgtgg	tattatgtga	ttttttttgg	246780
tgggggtagt	gtctggagct	gagcatgatg	gagcctagat	ggggcagcct	tctaagtctc	246840
ccagtatttc	gcaaatcata	tgtcacgtca	aaattgtaat	gacaaagctc	acagcagctg	246900
cctacgccag	gtttgtgtgc	agctctaatt	agagcattga	tcatttctct	ccccttcagt	246960
aatgcacatg	agtgacacatt	ttcatatata	agggcaaaag	acatttcaaa	gtttggatgg	247020
aatctttcct	acttagaagt	aataaaaaac	aatactttga	aatctttacca	aatgttaaac	247080

```

aaatagcaac aatttttcca tttgaagatg tgtgtcttaa ggtctgttgg atctttcaaa 247140
tttctatttt ggacatggaa actacagatg aaatttaact tactcaaagt tacacagaac 247200
tcctggctga tctctttttac ttaattcttc taaaatctaa agtaagaata attttgaaat 247260
tttattggta atgcttaaat aacaaaagtt actttaatat gttacatggt tgaagtttca 247320
tttatactta atatctaatt taacgcttgt gaaaatatta ttcagtattt ttttttcttt 247380
tctaaaatga cttgttcagg gctccttttg ggattttaag tgagaaaaac tctatttcat 247440
aatattgtta ccatgctcat ccaataacac ttcatagatg gttagccatc tctaaaatta 247500
aattttgtag ttgaggagca ttgaaaatat cacctatatt tggatttctg tttaacactt 247560
caaaatgcat tttatcataa tttaaaattt gattttaata caaacctctt tatagaaaac 247620
ctgataaatg tagttttatt tgtttgcttg cttttttttg cagggaggga tattcccttg 247680
ataccactct gtctttatga ctcatcgggt atttataaat gttatcatcg tttttctgat 247740
ttttaaataa gtttcagaat taatacattt aactgaataa atcaaccaat aatgtcaaac 247800
ttaaaatgtg tttgtaacaa gagtcagcac ttataatgaa aaatataatt aattaagagg 247860
ttayaaaact tataagcaga gtttaactctg cttacagaag ggcaaacatt tgttatttga 247920
tacaacatat taagtattta tcccatgcaa tcaactgtcct aggctttgag tttatagtga 247980
tgatcaaaac gaatttacag tttcaagaac cctgctcaca tacagcttac aatcctatgt 248040
gggaaaatga taatagtga ataatcat taattaatat atacttgatt caaatgtaac 248100
taattattct gaaggaaaga cacctacatc aagagagctg gcttaaattc ggggatcaag 248160
gaagcctttc caaataattg ttgatgggat tgaaagacaa ttgtgggatg agatgaccac 248220
ctgggcttgg tatgggggag atacaactgt cagtgggaaa gccctggaga tggggaagaa 248280
gaacttcttc aatggagaga agttcctgag tgcagkagaag gcaggtgacc acagtgtgaa 248340
gtgatgctat agaagccaac cagtggtgta tgcaagcaga aggaatttat ccaaaaagaa 248400
agggggaggc attagaggga tttagtgtgg agagtgttgg tgacaagatc agatttgcaa 248460
cttgcaactc tctctcaggt tgcagcatgg gggatgatgg gatgagcca gagggtgactg 248520
acgcatttcc agtagtccag gagagtggca tgtggcatga actgtgtgag gaagggggag 248580
ctggggggaa gaagtggggt aacaacaccc tatttgttgt tttgcttctt acactttgaa 248640
aaatatttca aagcaggaat gctattccag acatttaaca cctggaacag gaataagta 248700
caaccaataa tcgtgggaga ctgacgaatc agaatgaaca tcagctctgc atggcagtat 248760
ggttttaaaa cactttcatg tctatgtaaa caaagtaaaa atgttaaaag attattttac 248820
gcctttgggt atcagatagt cagactttta gtagtttata ttttattgca gtotgatgag 248880
gtctgtcttt aatagtagtt ttctcttctt cccccaaccc cattacatta aaatggcaa 248940
tgatactttt ccttggttagt aagcaaggcc aagttaggat tatcagaaag tcaataccta 249000
aacacaaaga atagtgcagc caaacccgga aaatattgtc tgttttgcca agaggaagca 249060
tctcactgag tatctggaaa gctcattggt tctcaattat ggttatttta ttagttgcca 249120
taactcagct atagtcataa tgttaactgac cagcaccgat gatgaaatca ataaccacta 249180
caacagagta aatgctgttt ggggtgcaag aaacatcatt tttatctaag tgaagtagca 249240
tctcatttaa agatatttca agttgacatt ttgtatatat tgttaataaa ctcaattttt 249300
actaaaatct ctgcaatata taatcaatat actgaatgat atgctctccc cttttgtgct 249360
caaataggat ctgtttatat tttagttaag aaatgctatg ccattttatt atattagaaa 249420
aatatttact ggtaggaatt tggtttaaac tctccctggt attcagctgg gaagtattat 249480
acaagtaagt ctttaaaagac tacaattgga gaccgggcac ggtggctcat gcctgtaatc 249540
ccagcacttt gggaggccaa agtgggtgga tcaggagttc gataccagcc tggccaacat 249600
ggtgaaaacc cgtctctac tgaaaataca aaaattagcc aggtgagggt gcaggcgcct 249660
gtaatcccag ctactcagga ggctgaggca ggagaatccc ttgaacctgg gaggcagagt 249720
ttgcagtggag ctgagattgc accattgcac tccagcctgg agaacaagag cgagatttctg 249780
tctcaanaat aaaaaaaaaa aaaaangaaa aaaaagact aaaattggag agtttaacgt 249840
atcaaattta ggagataaat ggtaataaat atgtttgtat atgaaaatat ttaatacaa 249900
tatttgaaat gaagaaactg agatcacaca actttttgag cgttgagttt atacatttat 249960
tttctttctt tctttctctt tctttcttct tttttttttt tttttgagat ggagtctcac 250020
tctgtcaccc aggtctggagt gcagtggcac agtctcagct cactgcaacc tctgtctcct 250080
aggctcaagt gattctcctg cctcagcctc ccgagttagt gggactacag gtgtgtgcca 250140
ccagcctggt ctagcttttt tgtatttttt tagtagagat agggtttcna ctgtgttagc 250200
caggatgggt ttgatctcct gacctcgtga tctgcccata tcggcctccc aaagtgtctgg 250260
gacgtgagcc actgcgcccc gcctattttc tttcttttac ctatatcatt tcattttttt 250320
attgatcagg aatattttac atatttatgg agtacatgtg atattttgtt acatgcatag 250380
aagggtgaat tatcatcagc gttttgggga tatatatccc cttcagttat tagcattttt 250440
atgttttaag aacattttcac attctctctt ctgcgtactt tgaaatatac aatacatgtt 250500
tgttaactat agtcactcta ttctgcatt gaacattaga acttaaaata ctttcagcc 250560
aggtgtgggt gctcacacct gtaatcccag cactttggga ggccgagggt ggccgggtcac 250620
ctgaggtcag gcgttcaaaa ccagcctggc caacatggtg aaatgtttgt atttgtaaaa 250680
ctacaaaaat cagctggggc tgggtgggtg cacctgtaat cctaactact caggaggctg 250740

```

aggcaggaga attgcttgaa cctgggagggt ggatgttgta gtgagcccag attgggtcat 250800  
 tgcactccag cctgggtgac aagagcaaaa gtctgtctca nagaaaaaaa aaaaaaaa 250860  
 gaacttatag ctccaacta actctatggt tgcaaccatt gatcatcttt tcttcattcc 250920  
 ccactctcag ccatacacac ttcccagcct ctggcatcta tcattctact gtctacctac 250980  
 atgagatcaa cttttttaat tcccatgtat gagtgataac atgtggtgtt tgtctttctg 251040  
 tgtctggott atttcactta acatagtgac ctccagttcc atttatgttg ctgcaaatga 251100  
 cataatttca atgntttttg tgcacaatta gtattccatt ttgcttatct atttgtaaat 251160  
 tgatggagat ataggttgat gctatatctt tgcaatgggt aatagtgcta cagtaaatat 251220  
 gccagtgtgt gtctcccttt ggaacaccga tttattttcc tttggataaa tacctggtag 251280  
 tcgaatttct ggatcatatg gtacttctat ttttaatttt gctcagaaat ctccatagta 251340  
 tttgttgata atggctgtac taatttacat ttccaccaac agtgataaaa agctcccttt 251400  
 tccctgtacc ctcaccagta tctgtcattt ttttgttttt ttttagtaata gccattctac 251460  
 tggggtgaagg tgatcacaca tagtggtttt gattcacaat tccctgatag ttagtgatgt 251520  
 tgagcatctt ttttatatac ttggtggtta tgccatttgt atgccttttt ttgagaaatg 251580  
 tctactcatg tctctcatg tcttttgctt actttttaat aggattattt gtgagatttg 251640  
 tgaggtgtgt tttgtttgtt ttgtttgttt gttgtttttc ttacagttga gctatttgag 251700  
 tccctgttat tttctggata ttagtcacct gttagtgaa tagcttgcaa atattttctc 251760  
 ccattcaaca ggtgttttg agattctctt taaaatagct gcaacaaaca aacaaacaaa 251820  
 caaacccggc caagcacggg ggctcatgcc tgtaatccca gtacttaggg aggctgaggc 251880  
 gggcggtatca cgacgtcggg agatagagac catcctagct aacacagtga aaccccgat 251940  
 ctactaaaaa tgcaaaaata ttgacctggc gtggtggcgg gtgctgttag tcccagctac 252000  
 tcgggaagct gaggcaggag aatggcgtga accgggagg cggagcttgc agtgagccga 252060  
 gattgtgcca ctgactgca gcctgggtga cagagcgaga ctccatccaa aaaaaaaa 252120  
 aaaaaaaac cctaggaata aatgtaacta aggaggtaaa atatacctac aagttgcaact 252180  
 aaaacacact gacaaaagaa attgaagagg gtgcaacaaa atggaagat atcctatgct 252240  
 catggatgaa gagacttaat atgattaaaa tgatcgtatc atccgaagca atccacacat 252300  
 tcaatgtaat ccctatcgaa ataacattat tttttcacag aaaaaaattt aaaaatcctg 252360  
 aagtttgtat agaaccaaaa aagagcgcaa atagccaaaa ctatactgag caaaaaggac 252420  
 aaagcttagt gcatcacact acctaacttc aaaatacaag gctatagtaa caaaaatagc 252480  
 atggtgttgg tataaaaaga cacatagaga ccaatggaac aaaacagaga atctagaaat 252540  
 aaaaccccat atttacagct agctgatttt tgacaaaaat gccagaaca tacattgggg 252600  
 aaaggatatc ctctttaaaa aatgtttgca ggaataattg atttccatat gcagaaaaat 252660  
 gaaactggac ccctacctct caccatacac acattcaac tcaacatgga ctaaaagact 252720  
 aaacgtaaga cctgaagcta taaaaccact agaaaaaac agggaagaca cttcaggact 252780  
 ttgttctagg caaacatttt atgggggaga cctcaaaagc acagaggaca aaaacaaaaa 252840  
 taaataaatg gggctgtatt aaactacaaa gctgcagagc aaaggaaagt attttcttaa 252900  
 tccttactgg ttcttaaaaa gttttggatg ttgaccaaca catctattat gataattaaa 252960  
 tgatttggac atgtcaaagt tgtcaggata tgcgcaggtt gccgcttta ggaagagcac 253020  
 agtgactctt tctaattcta tttttctatt tcttaaacat agctgactat aaaattgcaa 253080  
 agaaatttgt gtatgagcta gggaagatgt tgggttgggt tgagattcaa ggaaggagt 253140  
 ggactttaaa atggtggaat caccagtaga ggacaacacc gtgtaccaa aacgttcaca 253200  
 ggacaagcca ggtgtacac catgatatat gagttgtgat gctgatgaag gacatagtaa 253260  
 gcttccaata atgggtcaaa aaaaaaaaaa agaaaaaaag aaaagcatta atattaggca 253320  
 atgtattttg gtgcctataa tttatattct tatcaatacg gatagctcca tattgatctg 253380  
 tttatatata tctatttgtt tatatcattt gattaaaaata tggcattgac cattggaaag 253440  
 gtacatctca aattatcctg gttaaatgtg aaagtatgaa gaggtgtgta ttataagcat 253500  
 ttaagtttca atcaagtagc aagtgtgtgc aagatggtat atccaatcag atattgtgag 253560  
 ggacatgtga ttgctataga ttattctgct gtcttaagag ataggctgtg gatattcaga 253620  
 ttgtgatttt ggagttaaag agggcaaatg aaattgcctg ctgaactcac agtcattgtt 253680  
 gagagagact gacaggaaaa ggtaaaaatg aatcatcatc tacaactggc tggattggac 253740  
 ctgacggaat aagatccagc aggaagaggt tgtgactacc attcctggaa ctgggttttg 253800  
 ctgaatcatt ggacagtttt gctgaatcat tgagcataca agttttgagg gatacatgtg 253860  
 catttagaaa gtaacgaata agaattctg taattattct ctacttttca tagcccatag 253920  
 attaacacca aagcaatttt agatgagcta tcaataaaaa tattaaaaaa gaaatcacag 253980  
 tctattttatt taacacacat tggttatagt caacaaggta atgtgttgtg ctgcaataat 254040  
 acaagtaaac atccgtctaa gaatccattg tttagtaggt attactgtac gaatttttaa 254100  
 caaagctgct ttaaacacta taccaacagt tttaaataca aaagccaggt ggtcaagctc 254160  
 taatataaat gaataaatgt ggccttctgg aagaccctag agagacaaca ggacacagag 254220  
 tggagcacag gaagaaaaga agtccacttc ctgcataaat ggcggcgga actctgttcc 254280  
 acctgacaga cgctcctcag aaatgcagtc ccatttaaaa ctttttttga ttgttttaga 254340  
 taaaacttat aacattgttg aagaagcaac taattctaga tttttttata attttaagct 254400



gccagaaaaa	gtaagtccac	aattggaaat	ggcccaagge	ccccaggatg	taccttacct	254460
atgccaatat	tgagaagatt	attccagtg	aaaaggtcca	tgtaactttc	attaacacta	254520
tggtttgttg	gaaacacact	aaaccactgt	tccataattc	cttttttttt	ccagttactc	254580
aaagcaaaaa	taccagaatt	ttatgtttac	aaatgtgagt	ttttaataga	aacatgcaca	254640
gtgtatactt	tatatattgc	cggaattgac	ttatcgaatg	ataatttaac	aagattgttc	254700
atttgaagtt	tatgccact	tttataaata	aaagaatcat	tttggtgtgc	ttcctctctt	254760
tctctttatt	tttcttttct	tatttcagtc	tacaactact	cccccttctt	ctggcccatg	254820
tggagatata	ccatttccca	acctgctccc	tggtcagtta	caaggaggga	gggaagacat	254880
ctggcagagg	aagattccca	gggaactatg	aaaggagata	gtgactactg	cagagcttgt	254940
tacaataaca	gactattaag	gctggtgcag	tggtcacgc	ctataatccc	agcacttcag	255000
gaggccgagg	agtgtggatc	acctgaggtc	aggagttoga	gaccagcctg	gtcaacatgg	255060
tgaacccccg	tctctactaa	aaatacaaaa	attagccggg	tggtggtgtg	agcatctgta	255120
atcccagcta	cttgaggagc	tgaggcagga	gaatcacttg	aattcaggag	gcacaggttg	255180
cagtgcagcca	agattgtgcc	actgcactcc	agcctgggcg	acaagaggga	aactccgtct	255240
caaaaaaaaa	aaaaaaaaaa	accacaaaaa	acaaacaaca	acaacaacaa	aacagactat	255300
tagcactttc	tcatatccct	tgtttgctta	agaatagtat	ttgaataata	taacggccta	255360
tattttatct	acctactctc	tgacttcctc	tggtttggag	gggggtgaat	catcagaggt	255420
gttggtgcac	agtatgtgtg	tggttaaata	gggtcccttt	tctctactca	tcttttctat	255480
gcaaaataaaa	gccattggtt	ataatgaaga	ggaaatcaga	ggggaagagg	tctattttga	255540
atttttgtaa	ccgaaaaagc	tcattgatttt	atgtgaatat	tccattggct	gcttagcttg	255600
octacttact	tgaatttggg	gtttattccc	aacactgcag	cttactatga	ccaaggtttg	255660
ggaacattct	tgaatttttc	cttgaggaca	aatatccgct	ttgttgatt	ctttgttatt	255720
tatccatttt	gccaaattat	ctgcaagtag	aaatatcgaa	ataagaagct	ctttagcaat	255780
ttacttttga	tattggtttt	cttttgaagg	acagttatta	aaatagcttg	taggattact	255840
cattttcrtt	tttcttcttt	ttaaatataa	agcaatgtca	tcactttttt	ccctgtatta	255900
tattttctct	caataattga	tatgctacat	taaaggaaca	caaaatggct	ttaattatgc	255960
aataatgata	aaaggcaaga	gtgttctctg	ggaactaatg	gttgccctgag	aggaggtgat	256020
ggcttgaggt	ccagctggtt	attaagccgc	aggaaatgct	gcaggccaag	atttgtatta	256080
tttctctgag	atgaaaatga	acccaaaaaa	aggcaaaatg	ggtttttctc	cactaatggg	256140
taaaatgaac	tcatttttga	tgcaagaagc	ctttatgtag	aaggaaatga	ttccataaaa	256200
gcaaaatcaa	aatgttcagt	ggtctctgct	ttttgtttta	aatatagtat	agcacaaata	256260
atagtccatt	attgattttt	aatcttgatg	caactaaaaa	tatcatgcaa	cacttgagtt	256320
tttgtctagt	tctagaactt	ccatagaata	ttatgtccaa	caattgtttt	gatagctttt	256380
gtaagctaag	aatatatttt	gtatgcattt	aacattgaaa	atggttactt	ctgaaaatat	256440
tcagaaaaca	gctatctcat	aacaaaagaa	tacaggcatt	attttctttt	ttttctggag	256500
ttcattcaaa	aaatatggg	gaaaggtgtc	tcattaaaag	atgcgttgag	atttgtgtgc	256560
tctctgcagt	attagagatg	aatcacttaa	ccttcttaaa	gaaactgcct	aatcaatttg	256620
taaagtagtt	agtatagctg	cccccatctc	tgattttag	taggattatg	caaagcagag	256680
tattttgtag	atattaatat	tactaatata	aaaatgtagt	tattactata	tttactttac	256740
catcttaacc	aagtgaacaa	atatattttt	tcattttttg	ccattttttg	cacctctgtg	256800
acatggggta	tgtcacaggg	catagtagac	aactgaaaga	gtcctaata	gtcaaaactg	256860
gaaaagtttc	agaacaaaat	taaccagtag	tattagatta	taacccaaaa	tacaaaataa	256920
atgtttctga	attcattttg	atttaaatga	atgacagaat	acataagtca	atgggggaga	256980
gtttattctt	gcatgcagaa	gaatccagat	aatttatgtg	gatgccctgc	cttcaaagag	257040
gcagagccta	actctcattc	ctggacagca	cagagtgaat	tccttccaaa	gagtagagga	257100
ccgaaagggg	tgacccaaaag	attgaagaaa	gatcaagcag	ttatcatgag	ttaaagcaaac	257160
cttgacaagt	ttatcgcaat	gtttagaatt	ttgttatatt	caatogagga	ctttacaaaa	257220
ttgtgttaaa	ctttatgatg	gatattgggtg	tccttcttaa	tagagaaaag	tcaaaagccg	257280
gggtgtttgg	aggaaacatt	atcttttgtc	tgccactata	aatgtgacta	aaatagtacg	257340
taataagcat	atggccacaa	acttaaatgc	attatgtttg	caagaaaata	tggtggcagg	257400
tgtgtagcag	gccttgacga	agattttag	gaaaaatcat	aaaacaaagt	acatagtgtg	257460
agatgtttgc	tatacatgag	gattttttta	aaagcagatc	tagtagaata	tgattttact	257520
agagaagatt	tgtatctctt	gtgttcagct	gaattatagc	atctaacaca	gcctgtgccca	257580
cacagtagga	actcactcaa	tttttgttga	aattattcaa	cgaatatattt	attttaggct	257640
tttgcaagat	atgacttttg	taataaatgt	agcacactta	attcattgaa	cactaaagga	257700
aaacaaatgct	aaagaaacta	gtttccaggc	acaaatgact	tgcttaatat	cacaaggcac	257760
atggtggggg	ggggaggtaa	ctacaagatg	ccacctgaa	ggagtactg	tttcatctgg	257820
agcactagag	atagtgggtg	ctattgatgt	tatctagcaa	gactgagggg	ccaactttct	257880
gagttcaaac	cgtgggtttt	ccacctgtaa	atgaggaaca	acgataggac	tggcctcaga	257940
gataacccat	gaggactaag	ttgggttaatg	tagagtactt	agatagctgc	ctaaaactat	258000
gtaatcatta	tgtaaagata	agctattatt	aattttaatg	gtcataattt	aaatgaacat	258060



tctaggatat	agttgcagac	attgcaatac	accaaagtca	tttatgttat	aatttatatg	258120
ctatccagga	gagaaattgc	aacacagatg	tgcaggcacc	ccactatcac	cccaaattaa	258180
tattatgggt	attaattcaa	gacctataaa	atagcatgtg	tcattttataa	aactctcaaa	258240
tgtccatatt	taaaaacttg	cttaaaattc	ttaaaaatca	aaaagttaat	gggttttgaa	258300
gtctcttggt	aaaaaaataa	caaatcagtg	taaatgaatg	acagtttaat	atcaaaaaga	258360
tgaaaagtgt	atagccaagt	aacaaatgaa	attgttaagc	agggttgaaa	aattaccaag	258420
tgcaatggaa	aaattaacac	agggatttta	tatttgtaat	tattgattaa	cttcttctat	258480
taattcttat	tttgacaata	tttaaagcca	aattattggca	tatatataaac	tgagaaccga	258540
atccttaaat	tgctgtatct	cagggtacta	aacccaaatt	ttctaaaacg	ttgaaacaaa	258600
tttcaactac	ttgggcttta	ttgaagatat	tattgcttta	tccttttagtt	agaatgatgt	258660
tattataatg	ttgtttctga	agaacatacc	ataattaatt	aatgttaatt	ccccatttct	258720
tttatcattt	agatttctac	ttttagtgtg	tttcatgaag	gacatgaaac	attactattg	258780
aatgacatat	acaaaaagtt	gaagatgtgt	taatgttctt	ttattaggct	gatggaagta	258840
atcaaacata	tttattattt	atccttagtg	aagtttaaat	aataaagtct	tgtatactat	258900
taaactatat	attactttta	cttgtcacca	tcttttctct	aagcctgggt	tagaggtaaa	258960
aatcttttca	aattgtgtgg	cttcaacttc	caacttgctc	accttctctg	ggtaagagtt	259020
gaccataatc	tatcaaggac	acaaatactg	tattccttct	tcccagggca	tttatatagt	259080
cacagggaag	gccttagcct	gccttttgga	cttccctatg	ctttaagatt	tttttcttta	259140
tgtacgataa	gtagaagagg	aagaaataca	tcaaaagaag	tttcgctgca	attcaactac	259200
cttattgacc	ttaacttcct	caatcaacta	aaagtatctt	tctacatcat	tcatttcagc	259260
atttttccct	ttttcttatt	tatttgtttt	tcttttttct	tttctatcat	ccatttgact	259320
tttgtcacag	atagatttta	acagcttgca	tttgtatgtt	atfttctctaa	aatattagag	259380
cttctattaa	aagattgggtg	ttaaagaaatt	aggggtccag	gagggtcatg	cctataatcc	259440
cagaactttg	ggaggccgag	gtgggaggat	cacgaggcca	gaagattgag	accatccttg	259500
ctaaccagggt	gaaaccgcgt	ctctactaaa	tatagaaaaa	tttagccggg	cgtgggtggca	259560
cgcgccgtgt	gtcccagcta	ttccggaggc	tgaggcagga	gaatcgcttg	aaccaggagt	259620
tgcagtgagc	tgagattggg	ccactgcact	ccagcctggg	caacagagca	agactccatc	259680
tcaaaaaagt	aaataggata	atttatattct	tatatctatt	aacatatttt	gatggatccc	259740
ctgacatgaa	taagggcaga	gacggtagca	agttaagatg	ggccagttgg	atttcatgcc	259800
ttcaaggacc	tttcattcga	gcatagggaag	aaaaataaga	gtatacatat	ttataatata	259860
aggccaaatg	tttgaaatcg	tgaagccttc	cttctgtatc	tcaaagagca	atcaactctg	259920
gcacaagtga	tctacagtga	ctttgtggga	gcagttggca	tctgaaaaac	atcttaaaaa	259980
acagctctata	acttgctatt	gtttcttttc	acttacatgt	tatgaaagaa	gagaaataga	260040
gcctatgtct	gctatctcct	ggaccctccc	ctagtacatt	tttccattgc	taactttart	260100
ctacagcctt	tcaactgtaat	aaaccattac	taagagtgtg	ttagtgtttc	taaattcttt	260160
gagtcctttt	agctttcact	gaatctaata	tttgtgttag	gaccccccca	cacaccccta	260220
aaaaagaggc	tctagagttc	gtcagctggg	acattcaagt	ggccaggccc	aatatattca	260280
gaaatgctga	cttccttaca	aaaactaaaa	gctttataat	tttcaaaaaa	aaccccaaac	260340
ctagtttgaa	taacacataa	caagttttga	gtgaacgctt	agcactgtct	caccatgagc	260400
tcatggattg	gaggacagtg	gcaggggccac	acacgtgccc	tcctgttagc	caaagttagg	260460
ccagtttaca	gacatcagga	agttgccact	gcgatgcagt	cattcttgac	ctctgtgaag	260520
aaattcttcc	agttaaaaat	ttatttatgt	ctacaagcat	cctctttcaa	atgcaaccac	260580
cccgtggaat	attttaagtc	tcacagaaaa	gtcaccttta	ttttagcatg	aaatggctta	260640
aaaatcagat	tatttttctt	caaacatcag	atataatttc	cattaaggaa	atagaagtct	260700
ttatagtgat	tgatggagag	aagttatttt	ggcataaaaag	gaaaatgtgc	ttgggtggaa	260760
taccagagca	gacagctatg	agctgtcaga	gagaagaact	tgtggggagc	agaggagagg	260820
ggaccttgaa	agctgggaga	gaaaggcaca	aaaacagcaa	tgcaagaagt	atattttaac	260880
cactttcagt	ttcactacta	gctgccctta	cagagtgttt	gtgattaaag	aaactcttac	260940
taaccgattt	ttaaccgcgc	ctgagtcttc	catgtatctt	agcttgattt	tttttgccct	261000
caatcacgtg	tctacactag	atcaagtgcg	ccatctacca	aatcaaacac	tcgtttttta	261060
gtagggtgta	aactgtcact	aaggaatggt	tgggtcaatg	tccacctata	gatactgtct	261120
tcaactgtact	tgaggcagta	ttggagaact	tatgttttct	atggaagact	tagagcatta	261180
ctggaaaaga	gacggccttg	tttcttttca	tattaaatgt	caaggagagg	cggtaaagcc	261240
tctcctctgt	ttttctatag	aacccagttc	agggctgtgt	ctcatcactt	actatccctg	261300
gtattattaa	tgatcaattc	atctactcct	caccagactg	caaagttttt	gaggaaagaa	261360
ctccttcatt	ctccactatc	tacacagact	acgtataaat	taacaatgaa	taaatgaaca	261420
aatgccttct	tcatatatct	tgactagtct	atatgttgtt	agcattatat	ttaaatacca	261480
ctttttcata	agagttgttg	attagaaaaac	tgtgaaggga	ttgtttgttt	gtttgtttgt	261540
ttgtttgttt	tccttaaaaa	attaatactt	gcggcggggc	gcggcggggc	acgctctgaa	261600
tcccagcact	ttgggaggct	gaggcggggc	gatcacgagg	tcaggagatc	gagaccatcc	261660
tggttaacat	ggtgaaaccc	cgtctctact	aaaaatacaa	aaaaacatta	gccggcggtg	261720

gtggtgggca cctatagtcc cagctactcg ggaggctgag acaggagaat gacgtgaacc 261780  
 cgggaggttg aggttgcagt gagccgagat cagccactg cactccacc tgggtgacag 261840  
 accgagactc cgtctcacia caacaacaaa aaataataat aatacttga caggaagacc 261900  
 acagctatca cttctagatc taaacattct tcaagatgag tggagacaga ggtagagtgt 261960  
 gtgctggatg gatataggag tggacttact tggttgtcat catacggttg taccatttct 262020  
 ttttaataat gactgaagat gacattggga ataaataaac atagacctt gagtttcata 262080  
 gccacttaac tccctagttt tgtttttcaa attagttgag aggtgatact tatagccata 262140  
 actttaataa cattaccatc atcatctcaa ttgtggcagc tacaatgtca agaaactcac 262200  
 aaagttaaaa tgtgaagaaa ctttagaggt ggtttgggtt ggagaagatt gatgttgata 262260  
 cgattcagt atctttcagg attccatggg attgcactct gtcatcaatt tgctctaagg 262320  
 actaagtctg aaaaaactta ttgtgtgtca ttactgaaat actcagcagt ggtttgggtt 262380  
 ctgaatactt ttaaagcatt ctctctcttc tccgcattat accataaaat atatcttagc 262440  
 atattattgt tttcacttgc cttttgttta atatatgaga atataattaaa gcatatgcct 262500  
 tgaatcaaat gctaaccaga ttaaactata tctaagagag tgtgtctttt caaaacagat 262560  
 tgtgaagttg tgttctggtt cccatattac tccattgtt caatattttg ggacttctat 262620  
 tttggaaatg cattcagttc caactttag ccctaagaa atgctcccta catctaaagg 262680  
 ctaagatgcc ctcttttatt atttttctga tcaaatattt aagaaataag cttcaaatat 262740  
 ttgttaaat tttctcaaaa gtaaaatttt tatgtttaaa atatgacaat ctaattgtca 262800  
 tctttaatgt attaaaatct tttatatcca gtttttattt tattttattt acttatttat 262860  
 ttatttgaga tagggctctc ctctgtcacc caggctggaa tgcagtggca caatcttggg 262920  
 tcaactgcaac ctctgcctcc taggcccagg tgatccttcc accttagcct cccgagtaac 262980  
 tgggactaca ggcccacatc actgcgcctg gctaattttt atatttttag tagagacggg 263040  
 gtttagccat gttggccagg ctctgtctga actcctgacc tcagttgatc caccctctc 263100  
 agcctcccaa agtgcctggg ttacaggtgt gagccatata ttcagttttt agaagatcat 263160  
 tagaattcaa atgttgccag actttaagt tgggtcacta tctgatggag ggcttctctt 263220  
 agaagatgtc cccaatggca gggccttgtt ataaagggtt gaagagagat aatcctctga 263280  
 ctaagaggcc agagaagcag aagatttttt ttcatcagat ttcaagtcag taggaagaca 263340  
 gaaaggagaa ggaagagaag gcatgaacag catatgtgtg aagtagaatc ttaatagtgt 263400  
 gcatctttta agtcaggatt ccgtctactg ttggccaaaa taaatcatta ttatctgggtg 263460  
 gttggaaaaa ataaggacat tcttgggtt tctccaaaaa cttatttttg tgtttcttaa 263520  
 ttacccttag gagaccttaa tgatcaacac tgatcaaca gtattcacat aattatttac 263580  
 tcttatagaa catgtgatgtt attatgcaat gtattcacat aattatttac agttcttcag 263640  
 ttctagatga ctgatactac agaaacctga gataatccct gcttccatta gtttgggttg 263700  
 aacttttggt gagagacttg agtcatactg aataccttga attactgcca agagctgggtg 263760  
 atcatcaagc ttgagttaat atcaatgtta tgggtctatg gagaggataa ctgcaggaga 263820  
 acaatcatca tgtggtaggt ttaccacgtt ggcctgctga actccatgct caaactccaa 263880  
 ctacttttaa gtaatgacat taaaatcat attttccatt tatctgtgga agctagtga 263940  
 taaagaataa gatagagtct actgaaccta tagaatgtta ctaagtatta ggtgttgtat 264000  
 ttaatttcca gtttgtctaa tgctggaggc aattttacct tctcctacca gtgaaaaagt 264060  
 ttatgccagt aagacttgtg atttttgttc tcccccaaca catgcttota tataaattta 264120  
 ttatgggaaa agtgatatta ataaaaactt tctggaacaa ttatagatag agtgtggtct 264180  
 ggtgtgacag atattttgtt tcttcacaac attctctc tgagtattca ccattggata 264240  
 gatattctgt agtttctgta gtagatacaa aagctttaca aggtgtattt ttccacaaag 264300  
 aggttagcta gattaatagt gagaggatac atcaaaagac atggaaatga tttttagaga 264360  
 aatgagagca aataactaaag tattgaatta agtaaaacaa gaaaaatagc tctccacc 264420  
 ttcaaaaaa ttggtgagac caggattgtg gagggcacct tggcaatggg cctttagct 264480  
 cagtgaatgg ataaaatgtg tacacagtga gaagagaaga aagtgtctac acagaagtta 264540  
 gtttttgtt aatttgatat tttctcttta cctcgttcta gtcaagatga cctgaagcct 264600  
 tctagtgaag aaccagccag gcaagtcca cttcttaggg ccgttaaate ccttctctta 264660  
 ttaatgatgt taattgcatt tggatctcca tccagatatt tgccctacct taatttcatt 264720  
 ttaatgtgcc atgggctaaa gaggcagtat tttttaaacg catctgtaag tttcatttat 264780  
 gacagaattt ctgtaataaa agtgtagtca agagtgttga tggaaaaatta gggtagacata 264840  
 aatacgtgag aactggaaaa tgtataatct taacagtag acattatttt tattatctct 264900  
 ctttcccaca ggctgatgag gcctgggata ctaatatnaa aaacccaaat tagaaacttc 264960  
 tctctgtgg tgtttattgc ttgcctgca gtgcatgaca tcttaattct aaagctctna 265020  
 actcacaana tctgacatcg ttttacctc tgacacgtca aaatcagggc ctaattgntc 265080  
 ctgtggttgt gaattatgaa caaatatact tcttaagaag atggattttt catttaattt 265140  
 tctttttcag cttcacaata tgtacatata cagaggttgc ctgcgatttt aatatgtca 265200  
 cacctacaaa aacacacaat acacacaaac tttcaaactg aagcacacag atctgtaaac 265260  
 atacacatga agatttgtg caacacaatt cattctcaaa tctatttctt gttatgctta 265320  
 tgattaagat ctataacttc tcagggttga tttatttgtt ttgtacgaga ctgagtaaag 265380

ctaaagacca	cgaatatgat	gggtgttgct	aaaataaact	tctcaagaga	catcaagttt	265440
taatctgcc	tatttcaaat	attccctgtt	gaaatataaa	cataaaggaa	ccattaatta	265500
tataaaaaa	gaaatacatt	tgtgactttc	agagctgtca	cgaatagat	attggaaggc	265560
aggttattta	gaggactcat	gtaagacaga	aaagaactac	gcacagatca	cogatgatga	265620
cacagttatg	tggagaaagt	aatctgccac	aggtcaccca	gacagggagt	gggaaagtga	265680
tggctccgtc	ctcgatcact	tttatatccc	tttcattcca	aagccaagca	tgttctataa	265740
aagcagaaat	ctctgtactc	ttttgacatc	ctaactcctt	ctgctaccct	catttaccct	265800
catacatggt	ttatacagat	gtgggcttgt	cagaaaaatc	acaaaatttt	aatgttagaa	265860
tggacottat	catgctattt	tagttatcag	ctgctgagct	gttttctatg	tttgagttcc	265920
cttctctcta	aaaaatatac	aaaaatcaat	agaactttca	gtgcacttgc	agggtgggtt	265980
tacctttttt	aggccttgat	ctcacagcct	gatccttaac	tcactctactt	tccaatactg	266040
gctgacagag	catccacccat	ccaaaggcag	cgtagacaaa	ggttcaatga	gtaggaaactc	266100
ttaattatga	cataatgaca	ttaataaactt	ttaccttatg	ttttatgttt	ttggagacaa	266160
aataattcca	aacttattat	tataaccata	aacataaata	tttgtatcat	taataacagt	266220
gtaggacctt	aatttatagt	ttaaccattc	acatttgagt	ctttaatccc	tgctcaattc	266280
caaagtata	catagtttta	ggggtccatt	tttattttctc	tccacatgta	gtgattttatg	266340
tttctctgta	ctctctacta	agtaacctgg	ccttattcca	ttatatattt	agtttccata	266400
tatacataca	tctgtctctg	aggactttta	atttcatttc	attggtcaat	gtatctgttc	266460
ttgctgttgt	aacagactac	tgctactaat	actactgctg	aattactatt	attattatta	266520
ttacttttca	gtattttotaa	agagcttttt	aaaaatatgt	cattaagtta	ctgttttaag	266580
taaaacattt	tatatagaga	ttcatatata	tattgaactt	atgtttttat	tttcttgcat	266640
gtactttaag	tgatcttgaa	atcaacatta	taacaacctg	ataaaataga	gtgaaggttt	266700
ctgatctttg	ctcattttgca	ggagcacctt	atataagatg	taaattaggt	aaacgtttta	266760
catttgaggag	aaattcacta	tacaacaatc	tgattctgag	actccaaaga	togtttttta	266820
tgtttaaaat	atctattttct	ctattcaaac	tttcatttta	tttgccgcat	ctaatatctg	266880
acacatttcc	aaataagtaa	aaacaaaatt	caccaaatct	tgtatctttt	aaaatgtatc	266940
tttgtttcc	gtcttctatg	ttttgattat	ttcttttaaa	tttttaaaat	aaatttgtgt	267000
ttagttttga	accataaaaa	taactcattc	aaagtcatca	aagtaaaagt	aatcatctct	267060
taggcatata	gttgtcagta	atctgatttt	aacccaacag	actatatata	tatttaagac	267120
agggtctctc	tctgttgccc	aggctggagt	gcaatggcac	aatcatggct	cactacagcc	267180
ttaaactccc	aggctcaagc	gatcgtccca	agtagctggg	gactacaggt	gcacgtctcc	267240
actcctggca	aatttttgta	ttttttgtag	atgtgggggt	ttgccataag	acgaattgtt	267300
gagtcaagat	gtagatggca	cattagtcca	cgttatttca	gttaagggtt	cctgagcctc	267360
tactttgtgc	catgatataa	gcaagggtgc	aagtctgtgc	tgcaaacagc	agtcacctca	267420
tccaaaccgt	acacttacaa	gtgaggcaac	aaaggatat	aggggttcgg	gaaaaactac	267480
ggaaagcgtg	tgtgtgtgtg	tgtgtgtgtg	tgtgtgtatc	aggcagaaaa	ttatcctcac	267540
attttccctac	gagtatccca	gttgatccctc	tgccaccatg	ctctttccat	gggtgcccct	267600
tgctgtgtgc	tttccacccc	acccttccag	agggatctgg	atgtggatgc	ctatgctgag	267660
tgtaagtgaag	ccttgagatgt	tgtagaata	taataaata	ggccaagtat	gctaacattt	267720
tttcaatcca	caaaacgaaa	attagatctt	gacctacaca	aaatgtggct	agctcctctg	267780
ttgagaaacc	ctgacaccag	atctattaat	acctttccca	tggtcacttt	atgttttgaa	267840
agtttgctac	tatttggtga	tgacaccacc	aactaggatg	aatataaatg	ttcattttca	267900
aagatgttaa	aggcatcatt	ctcttatttc	taataattaa	ataataaatg	tggagataca	267960
cacacacaca	cacacggttt	ccatagtttt	tttttcaggc	aaattcgcta	aaatctcata	268020
aacattataa	ttattattaa	ctcctaggaa	taatatctac	atttaaggaa	agagtaattt	268080
taatttaagt	agacacattt	aatgatgcaa	gtatggatac	ctcatgcaaa	caaatgcctt	268140
tatacttact	cagaataaat	aatgggaatt	ccacatggaa	gcttatgaga	aggctcttta	268200
tgatcttatt	ttgcttttgt	tttagctttc	tgagtaaaca	agcccctaca	cactgagtat	268260
tagattcacc	tggaggtott	atttcataat	gtatccaggg	ctttagccct	agacatgctg	268320
attaagactg	gggtgggctc	aagaatttgc	atttttcaca	aattccaagg	tgatattgat	268380
tttattotta	gtagatcttc	tcattaaaaa	taagtgtaac	cctctgcaat	taacagaata	268440
aagtcaaaat	ctctatcgca	ggattcaaca	gagggatgac	aatctggccc	aaccaccttg	268500
catagctttg	ggacctaaca	ttactcttgt	tttacagcag	tagttctgcc	attcgttaag	268560
gaaatgtggg	ctgttctctg	aataaccat	tctcttttgt	atctttgcac	tttgatcttg	268620
ctgtattttc	tggccagaaa	gcccttcac	ctttacagt	tagaaactgg	ccacttacc	268680
ttcaaggcca	agttccaata	aagtttcttc	taagaagagt	gctgcctatt	tccaaattta	268740
agtatataaa	attgaggatt	gaggaaatcc	caaaattcca	gtacatttgt	ttcataatgc	268800
aatctaagat	cattctgtta	atgttccaag	tgcttccaac	ggctccaaga	agtggtatga	268860
gactcaagtt	gaattttgaa	tcacaaagt	tattcattgt	ccacctgaca	tcttttttta	268920
ttttatacca	tgttttttat	gtaataacac	atttgtgtct	ctaagagttt	ttcacgtgga	268980
ttttaataat	gtgccaattc	acaaaagaat	ttgacttaac	acttctgagc	attattcatt	269040

gatacacagt	tgaataacag	atagtttaga	taggaggaaa	agtcagtata	aaggaatgaa	269100
atTTTTTTT	aagctcaaca	tttcttagag	attttacttc	tcataatctaa	tgggtgaatt	269160
ctctcttatt	aaaaaatgtg	tttaatttaa	ctattataat	tttatcagat	gctgatttcg	269220
tgtgtgtgtt	ctgtgtgtgt	ttgtgtgtgt	gtgagagaga	gagatgatct	gatgtgtatt	269280
cctaagttta	caagattcta	aaccttttca	cttataaata	gatagttcag	aggggtgacat	269340
ttttggtaga	gaaatcttaa	tgaagcttaa	aatgcaagtt	tctgtgtaga	ggtaccatct	269400
ggttaaatctg	cttcaaatga	tgggtggatgc	accatcatta	cccaaattggg	tgaagcctg	269460
gctgcctata	gtctccctta	aaggttaagg	agtatatat	gggtgctaaa	gaaaggctta	269520
tcaggctact	gtacatcaga	gcactgacaa	aacttcctca	ggaaatcacc	aaatgaaaca	269580
aggacctatt	cagaaatggt	ggtttataga	tatgccatta	caatggaact	catctaaatg	269640
acagcatggc	ttaagtggag	aatttttact	cccaggctcag	gaaattctaa	ggcaggaaat	269700
aatgcctatc	ttaactcttg	agatgtggcc	agaattcagg	ttgtttttct	actgtgtgct	269760
ggttttctca	gacatgtgaa	tttggggaga	taagaacaga	atagcatatt	agctgcaaca	269820
tgtggcactt	gtcagattaa	caaaaggagt	gggaggcttc	atacaagtct	ccattttccc	269880
tgatagagtt	tcaagttgag	aacagtataa	tgttcatatt	ttacatcctg	atgggaaaaat	269940
gggtaagaat	aaaagggtgtg	tgagccatag	aaacagaatc	tgtttctttt	cactttttat	270000
agcaatagga	gtcaaaaggca	ctctaaaagc	tcacttgaag	gccacaaggc	cttgcctttg	270060
tcttttggtt	tcatttttaac	tttctgcttt	tttacaacaaa	aatctatatt	ttgtcccatt	270120
atttaattggc	aagttgcctt	taaaatgtat	ctggcctcat	cacagaaata	ctaaggtag	270180
tgattttgca	taaaataatca	cagagaatgg	catttcttaa	gttaattgtg	tctttctcac	270240
taattcccat	caagagcgtg	aaagctacaa	tttaggtgat	ctttcagacg	attagtgtag	270300
tcatccagga	agtcctcttt	gttatctaata	ctaaagtcact	ctagctataa	ttccccttgg	270360
tttttttttt	tttttttttt	tttttttttt	gagatggagt	cttactctgt	tgccctaggct	270420
ggagtgcagt	ggtgtaatct	tggctcactg	caacctccac	ctcccgggtt	caaacaattc	270480
tcttgccctca	gocctcttgag	tagctggatt	acatacacgt	gccaccatac	tcgggtactt	270540
tttttgtatt	tttagtagag	acgggggttc	accatgttgg	tcaggctggg	ctcaaaactcc	270600
tgacctcata	atccaccacc	ccctcagcc	cccaaagtg	ctgcgattac	aggcatgagc	270660
cgtgtgtccc	agcccttgtt	ttgttaggtg	agaaaatgag	cctgtttcca	catggagggc	270720
agtgggtggtg	aagttgtatt	atgttttgaa	aaatactgaa	agcaaacagg	gagggaagca	270780
gattaacatg	tttttaatca	tttatgcact	aatattgtat	acaatgtgtg	tgtgtgcctg	270840
tgtaaagaaa	gaatacactg	catgtaacac	agtaaggat	tgtatttata	tattgctcag	270900
gcagaagata	tttaaacagc	ctggattaat	tcactcatct	tgtttgtact	aagaaacaaa	270960
gactgaatga	gttacaataa	aacaagttcc	ccaaattgta	ttatttgtgt	ctcatttgag	271020
tgctcataa	ttgtatatgt	attgcattta	tcacaaaagt	tatatattgaa	cacaagggtg	271080
acagttataa	cgtgcattgag	caaaactata	gtgagccata	tctagcgatt	tagctttcca	271140
ggtgaaacag	ccactctagt	tatgtaaaat	gataggaaat	aaatcaatcc	catcttagcc	271200
taggacgaca	ttgacatcag	catcagttat	tccaataact	gcataatcaa	tatttatggg	271260
cttaatcaag	agagacatga	cacctggatt	cgttgcattg	gagaaggag	ggcaatgtca	271320
aggtggattt	ggtgaagtcc	aggtggggaa	agcagagtaa	atcagcagca	aagtcaacaa	271380
gcaatctgct	cacagatcca	caatattcgg	ctgctttata	gaacacagct	ttattctttg	271440
tctgtgatgg	aatatcaatg	cttccatttt	gggaaaagtc	taacgaattc	tttagctttt	271500
taaaattaaa	aaaatatata	tatagccagt	ttgtattttg	tattttttat	aaaaggcata	271560
taaaaatgca	aagcatagtt	tctgtctttg	agaagttttg	aatcttgttg	gtgagtcaag	271620
aaatgtagaa	gaagcaatta	taagcaggat	ttgaatccta	cctgtatcac	taatttgttg	271680
tggtgtacaa	atcaattctt	gtgcctaggc	caaaataaag	caaaaaaaga	ctacactggc	271740
tccacattcc	aagtaaat	actaattcaa	tggtactata	aaatgtcaat	gaatatattc	271800
cagaaaataa	gagtgaataa	tttcaagaat	gaaggaaaag	aagagtcaag	tgtggtgaaa	271860
gaaggcttgg	attagcaagc	atttgattga	tattaaaagc	aaggataaga	atttgctgaa	271920
tggagagggtg	gggaggagtg	agggttatta	tagaaaaaaa	atgaagagaa	taggaaaaca	271980
aacacaggtg	aaaaatgagc	atgaggggat	cacctgggat	tgagaagatg	acatcaactg	272040
ctggaaatgg	aagttttgca	tggagagcat	caggtcagtt	tacagaacgc	catggacact	272100
gaaataacag	ggctgggttt	gaagctgttg	aaaatttcaa	tcagggaatcc	tttccatctt	272160
gactcagaga	aaaggggtgt	ttggaaaatg	ttcatctgaa	gcaggggggc	ggggcgggga	272220
gcaagattaa	actggtagct	atcatttcca	gatattgccc	agtcctgaga	atgtttgatt	272280
attaaacaaa	accaaatgct	aacatcagca	ggtccacaac	aaacagaaga	tttacaatct	272340
cctgaaattg	caaagatctg	gtagaataga	tttcacatgg	caatagttgg	cagaagctaa	272400
gttgcttctg	ctttccatgg	ctctccttat	ctccacattg	ttgtgactga	agctggcctg	272460
ctccccattg	tcagcctgga	acagaaaagt	gtttgcatca	gagatcacta	ctgtagggag	272520
catagattgga	caagtgcctt	aggaaactac	taccgaggtc	tcacccctgag	atgagaaggc	272580
tctcatttag	aacaaacatg	gaaagggaag	attgtatgtg	agttttgttt	tgaagcaga	272640
atcaatatac	taataattga	gactgagtga	actagtctag	gaggcaagg	agaggaacta	272700

gcttacacac gaggaagctt aaaaatattg tttgtgtata ttaaaaaataa aataaatgcc 272760  
taggcatggg ggctcacacc tgtaatttca gcattgtggg aggccaaaggc gggagaatcg 272820  
cttgaggcca ggagtttgag accaacgtgg gtaacatgga gagactccat ctctacaaaa 272880  
aattttaaag ttagcccaat gtgggtgtac ttacctataa ttctagccta ctagagagtc 272940  
taagatggga atcacttttt tctgtgagtc taagactcca gtgagctgtg actctgccac 273000  
tgtactccag cctggatgac aaagtgaac cctgtcttac tccaaggaag aaaatgaaaa 273060  
taatactttt gaaagtcctc agatagaaga ctcttacctg cttcatttta ttagtataat 273120  
atatgacaaa ttaaaaatat taaatttaag tgtttatttg ccacttactc aaaattttag 273180  
gtatacacag gttgaccag aaaagaaaca aatgctaaat ccaaaaaatat tgtcgtgctt 273240  
catgcctaca ttacttttga gtacacatat ttcttctgta acccattata aggagttttc 273300  
aaaatataat atgttcttaa taacttacat taaataattc agtggtaaat acaagttata 273360  
atataattgaa tgttgactgg tcaagagttc ccttttttca taaaacagat ttcaattctc 273420  
atgttagttt ggccaaactga aagtaaatat gcttggtaga aaacagtcta agtgagacat 273480  
tattaatgaa ataatatgtc tgaaggtctg gtactaatgc tggatcacga atataatttt 273540  
tgaacatgaa gagctactaa atgtaactga gtttagactg ggaaactttt ccagtgatcc 273600  
actaggtact agtttagttg catggctgaa tgaccagctc atttgccaga ggagaggga 273660  
atgatatgta gctgactgtc aaatcatatg ctatttatta tagttatatt ttgtttgttt 273720  
taactcttaa gtgaattgtg agtttctgca aatcagaaaa ttgtttatta atgtttgttt 273780  
gtactgctgt gcttaattca aaaccacctc atgagtgatt tttgaaatca acccaattca 273840  
tacatgatgg gtagatttaa aaaatagttt tacattgttt cttatttctg ttagaactgc 273900  
aatcccga taggaagatt gcttcaaaaa ctataacaga tttctgaaat tgagtaaat 273960  
tgcttcaatt gtctattaaa agttttgaag tataataaac aaaaaaatat attagtaaa 274020  
gcaaggttca aaatcacgtt tccagtaaac ttgtcatttg cagtacattt atgtagacaa 274080  
aacaagtcca caatagttta aacagacatt cacagcaggt gattccaaga ccattgggaag 274140  
gatttggttc ctgtttattt tctgagttgt cttggatgct ctccttccag ttatgtcttc 274200  
ttctctttct catgtatcat gtcttatttc aactattctg acagcatgtc aagcattagc 274260  
agcagaataa tctgtatgca cagtgggaga gagaggcaga ccaaatacac attgaataag 274320  
gaaaaagaag tttccagcgg ttgatacatt attctcccta gcaaagggt agatttgcta 274380  
agcatatgga gacatgctta aggatcaata gggagaactt tgaaatatat gccatcctg 274440  
tgaaaattta tatcttgact gaggatatgt tttgatagta taagtaatat tctcataaaa 274500  
tttttgaggt gacacatgag aaatggagga aaccataatg taaaagcac atgttaagtc 274560  
cttgccaatg attaagccat ccttttgcct tataaagaca ttgattttca ttctatcctg 274620  
atcttttaga gtagttgggn ttctatatat cttgagaatt cctatgagaa taaaacacag 274680  
ctaaatttag tgtaattaga ttaagttaga ggaagcaag ataccctgga tcttgagtaa 274740  
atatacttgg ttaacgaat ttcttaagat atatccatga ggatattcatt tctcttacca 274800  
aaatataatg catatcttta tttatttgtt tttttaacct cagagcttct agggtagact 274860  
caaaaaagac attctgctag ccttatgttt tactatgttt ccataaatgc taacatttct 274920  
ttcaaatatt taccoccaa gcttatttta gatgtatcag ctaagggaga ggagtcagaa 274980  
agatacctta taatgaagtt tctcttctt tttcttatat tcgtaacaat ccgataaatg 275040  
gtttctgaat tatcgagtag acctctgaga actcagcagg tgattccaaa tcaaattoac 275100  
aaccttccaa gatgacattt tccctagaaa atcccagtat ctaaccttac caactggatt 275160  
aaaatcaaac ttctttatga aaaaatatac atcattaaat tgtgaaatca aatatagtc 275220  
tattaattaa aatgtaatgg taggaatagg aaaaagtgtt agtgtgcgtt gtaatagaaa 275280  
gagatgacaa gtgogcacag ctgaggaacg ttccttatat ttccaaggtc tctgtgtcag 275340  
aagttcgata tcaccataac cttttgatgg gtcttagtta ataaatagtt cgttattata 275400  
ttttattcat actgtaatat tgctttttaa ttcgtcctct tcattcatgg ctgtctttaa 275460  
atcaatagaa tcagttaaaa tatacagaat ataaagtaga aaatacagcg tattttctgg 275520  
ttagaatatt atgttgattt aagacttaat tgcattctgt gaaatttagg tttaatcatt 275580  
ccacaaaaaa acacttttag aagcaaaagt gatgattaaa gtattcaaca aacataattt 275640  
aataatattg gtttagttga caaaataaaa tgttgtgtg aaagccttcc ttacccttgc 275700  
tctaaaaacg gaaaaaagg ctgggcatgg tggctcactc ctgtaatccc agaacttgg 275760  
gaggttgagg tgggtggagg atcccttgag gtcaggagtt tgagaccagc gtggccaata 275820  
tggcgaaact ccatctctac taaaaataca aaaaaattag ccagggttag tgggtgtg 275880  
ctgtagttcc agctgctttg gaggtgagg catgagaatc acttgaaccc aggaggcaga 275940  
tgttgcagtg agccagatt gtgccactgc actccagcct gggcaacagg gaaaataaga 276000  
acaaatgtaa tctttagttg cctatagttg tttattacac aagcatgtgg 276060  
tgatttatat ttaatcataa tttattgatt atattaaaa atagagacac aatgggtctc 276120  
cagagaggga atgaacatat attattttgg ctaactatat tttatgagta actgcattta 276180  
catttctctg gttttgaact tagccatgct atgaacaact gaataatctt tattgttata 276240  
gtttcaaaat atttgaaatt atctaattgg tcttagatta tacaataaaa attataaaac 276300  
attatattga aaaaatgtat taatattaca aaaccttaca ataaaacata ataatacagt 276360

gttggttttg	catgtgtctc	ttttatataa	aaatcagaag	gagaaaaatcc	agttattttgt	276420
tctttttttg	agatttttact	aaattacaat	gccctcaata	atatgaagac	atacaattta	276480
ccaacaagtg	gccgggagct	ctgtctcaag	cctgtaatcc	cagaactttg	ggagccgagg	276540
tgcgtggatc	acgaggtaaa	gaagtgaaga	ccatcctggc	caacatggtg	aaaccttgtc	276600
tctactaaaa	atacaaaaaa	ttagccgggc	gtggtggcgg	gagcctgtag	ttccagctac	276660
tccggaggct	gaggcaggag	aatcacttga	acctgggagg	cggaagttgc	agtgagccga	276720
gatcgaccca	ctcactgcac	tccagcctgg	gagacagagc	gagactccgt	ttcaaaaaaa	276780
aaaaaaaaaa	aaattacca	caagttgata	atttcttaca	gaattaacgt	caccctgaag	276840
tttacccttt	tggtattaca	gctgtttatc	atgtacataa	ctcccttttt	attttgagag	276900
gggttaaaagt	taattcgtat	ctataattta	tagagtaaaa	ttatattttgt	taacaaaaata	276960
aaaacttttt	ataatacttt	aatgaaaagg	ttattcacca	aagaaaaata	taaaaaattac	277020
agtattctcat	caataatttg	gcagatatct	caactcttat	tttttctcgt	atttttaagac	277080
tataaattat	ataaaagtaa	cagtattcat	tgctaaacag	taaaacaatt	tttttctgtg	277140
tttacatcta	gcttactgtt	ttagtaataa	aaatgttctc	aattttcatg	cattgcagtc	277200
acatccttaa	aaaatctgag	atttcttgat	tttctgtaaa	tttggcattt	tagtctgaat	277260
tattaaactt	aagtaaccag	gatgtaaaat	tttatcgctt	ttttttcatt	taaaatttaag	277320
aatgacatct	ctctgagttt	ttatgtccaa	aattctggat	acttctaaaa	ggatgtgagc	277380
attactgcag	ggtagatggg	aatcctagac	tttggaatac	aaattaaatg	agctgttcta	277440
tctgaccatt	ctctgtaaca	cagtttcagg	gacttgaacg	acctgatttg	aaaaacaagt	277500
atgaacagtt	agacataatg	tgatctaaag	gtagagtatt	tatgtataat	cttaagttac	277560
aaaatttatt	gctgaaaaact	ttaaagtcat	atcgatatct	aatatatata	accttattta	277620
tatatattat	atttagtcat	ctgcctttct	ttccaaaaat	gctgaaaaat	aagtcagtta	277680
tcttttctct	catttgggga	ccatgagaat	tttctctttg	agaaatatgt	gttttttttt	277740
ttttgttttg	ttttgttttt	ttttttttgn	ggcagctctg	tgagctcaaa	tttcttaggt	277800
ttaatctaag	aaattccaag	agcgttactc	tcaaccatga	aagctttttg	ataggagtga	277860
actgcagatg	atctggcttt	ggaacattac	cagtaacttt	caaagaattt	tcttcatgaa	277920
agtcagaatg	tcagtgtgtg	tgctaaaatt	tcctggttaa	acatgtttct	ttttcttttt	277980
tttttttccc	cagctgaaaa	tacaaggcaa	acactgcgtg	tcaaaaaacat	taaaaaatcct	278040
tgcttttatt	ctcatattag	ggttcaatat	cagtggatgg	aagtaaaata	gtgtacacat	278100
cactcaaaat	aatcaatata	ctgagagctg	gtagccttac	tttgagttct	cacttctgct	278160
tttatgtaag	ccttgttgcc	tatttggaa	gtttttctac	ctgatacagg	aaatgagaag	278220
gtgatattcc	ttccagttgt	aatcatccac	gtttctatgg	tgtaaaagtca	ctgtattcct	278280
tttaccoccca	acttttacta	cgcatacaact	gcattgccagt	tactggctga	gttcttcaca	278340
ttaacaatga	cattaaaaatt	tcaaaatacg	catttgagat	agggataaaa	atggcagaga	278400
aatcaagcaa	gttcccaaag	gtcacagagc	ccaataagaa	acacaaacag	aataacatag	278460
gagtcagatc	agtagccaag	atttgatttg	tgtgatcacc	ataatataga	tcttaaacca	278520
aaactgatgg	ctacttaact	taattttgtg	aaaagaatga	acacagagag	aaagatgggc	278580
ctcatacaca	cacacacgca	cacgcacaca	cacagcatat	ttaactgcat	tctactaaac	278640
agatactctt	tcagggtgaca	attggctagc	cttttttaaaa	ttacttaatt	aattattatt	278700
tcttttttaag	cagagtcttt	ctctgaaacc	caggctggag	tgcggtggcg	ggatccggac	278760
tcactgcaac	ctccaccccc	ggggttcaag	cgattctcat	acctcagcct	cccaggtaac	278820
tggaatata	ggcatgtgcc	accatactgt	gctaattttt	tcagtaaaaga	agaggtttca	278880
ccatcttggc	caggctgggt	tcaaaactcct	ggcctcaagt	aatccgcttg	ccttggcctc	278940
ccgaagtact	gggattaaag	ccttgagcca	ccactccagg	actggctggc	ctttcttgat	279000
agaacagtta	agtttttaaaa	gcaaatgaga	aatactctct	tttttaaaact	ttcccttttg	279060
gatgaaattg	tgaaattata	gctaactacc	attgactaga	caaaactcca	actgtgtttt	279120
tggtcttgca	tttggttatat	ttaatacttg	actcaagaaa	accttgtaaa	gtgtcttttg	279180
ttattacaga	atctagaagc	cctaagcaaa	tatatcactc	tttttctata	tattaggaga	279240
atccccattt	gccctatctg	tattttcaaa	ctttggggct	tcaaatgttt	gaagtaaaat	279300
aatattttgta	ttttaaatatt	tacctagtca	attttctaa	aaattcaact	tactaaatgc	279360
ttaatgtcaa	tgctattcaa	acacaaagta	tataactac	acttaattaa	ggattcatat	279420
acagagagtc	atggaaaaata	taactcataa	tttataattt	acaagtctca	acattttgca	279480
tatttttcaa	aagggaataa	tatatacatg	tatacatata	tacacacaca	tttagatatg	279540
aggtaactaaa	ataaatataa	tgtaacttac	taataaaaatc	aatgtattgc	atgaattgtc	279600
aaaattctat	atttttgcta	tactaattca	actcttttgt	gccacattat	gttagagtca	279660
cttattttctt	tatggagctg	tcagagctag	tttactcagt	aaaatactct	gtatgtttac	279720
agcctttata	tgatattata	cttaaatcat	atgcaatata	attcatatta	tgtgcaatgt	279780
aatccaaaayg	tattctttga	aacatttgta	atgtgtattc	agtaaaaaata	tactgaaata	279840
aaaggttaggc	acctggcaca	atccatctct	tagcccttct	cacgttgtat	tgaaatgatt	279900
tctatccaag	tttccattgac	cttttagtcta	agctcctatg	gggaatttca	taccatgctt	279960
ctacaactaa	aatttaaccc	catactgata	aataatttgt	ggtcaataaa	tatttgtttc	280020

agctggctga agaagtaagc ctttccaatt aataataaaa attaaaattt tatgggattt 280080  
 tccaaataat tagaatcaga ggaagagcga gagagagaga gagagactga gatctcaggc 280140  
 atatcaacct tgtggataaa tagagggtta tagaaatgaa ttgacaaacg tgctgtaaat 280200  
 tgcttagctc tgcgaatctg ctaattctca ggagagtaga aaagacttga catttatcaa 280260  
 gtgtgtgtca tgtgaaacac ttcgtgcaag gtaattttca ctctttattc ctttttagtac 280320  
 caataggtat cagatatcac tgcacagtta gctgcaccta attatccagc taatttttct 280380  
 acttatactt tgaatacagt cactatatga caatggtaaa ctgtagcatt ttcaaatttt 280440  
 accgtatgac ggttgtaaaa atgaagttaa ctacttttaa acattccatt ttaattttata 280500  
 tatatatatt tgttttctga tttttattat cgcataatgaa tgattttata tgtgattaat 280560  
 ttatataaca aagtacacca ttgaacttca tgaatatgaa atcataatct ggtatacata 280620  
 gacttttagt aatataacag catttttagga ggaattatat tatgagctga gtggcaagga 280680  
 taacaaattc tgtcctgttt taattattta agtcacagtt ttactagtag tagaaaacat 280740  
 tctgaaaaaa aaaaaagaaa aagaaaaaga aaaagaaaaa caaaagatct ggtcagatgt 280800  
 gatggttcat gctgcaatt ctaataccct gggaagctaa ggttgagggt tgtagaattg 280860  
 ctgaagacca ggagttggac ctggcgact tagtgagacc cctctctac aaaaaataaa 280920  
 aataaatttt aatttgaagc aatggtctga gaaagtgaga aattgtcgac ttgtaccctg 280980  
 gataatttta tgcctgattt tgatttttct aacattttct actttaaaaa ttatccttga 281040  
 aaatttcaga agtggcttca cacttaattt tcatgatcat tttcaatgtt tcttccaaaa 281100  
 aatgatctac atggtaatta ttttatgtgg cataattctc tagttgagga gacacgatgc 281160  
 ttttactccc caaataaaaat atggatatat agctctttgt catatgagca catgttgcaa 281220  
 gaatcatttg atatgtttat gctgaaatag aaccttaatg actcatttat atttccttac 281280  
 atccaatttt attctgcatt acgaagtgtt taaataatat ggtataatat attgtttgag 281340  
 tctagcctct aacatcctgc aaactggcat cagaataatt tacttttctt aagattttaa 281400  
 agtcacaaaa atattcatta aaattaaaaa cattacatct ggaaatacaa ctttgctgtt 281460  
 taaactgatc attgtactct actttccttt tttgtaatcc gaatggtaag gtataattta 281520  
 caagttatat tttataagtt cataaatgag gagtaaatth tgttttataa aaacaaacaa 281580  
 taattgttaa gttagtattt atcacagggg gaaaggcctt gattcaacaa tccctgccaa 281640  
 gttgcaagtc tcaaatccta cttagatata tattacttgt ttgcgtactt ttaaatgtgc 281700  
 cttccttcac tacagagtat atatgctgat ggctttattt tttagcactt acattcatag 281760  
 attcctgcca tatgattgca ttttacagta taatacgaac atgttcataa agatttcatt 281820  
 tgaatctatt cacagttttt atatatactc catgaaaatg ttatactttt aatatcaatt 281880  
 atgtgatatt ctcaaattgt cattgattat tctggtccat tgttgagaga tgttcctcac 281940  
 tcttagtaac agtttctaca tcttggaatt cataataata acaagagaca tttagatagt 282000  
 gtgatggtta atagtgtcaa cttgattgaa gtatgcaaag tattgttctt ggggtgtgtct 282060  
 gtgaggggtgt tgccaaagga gattaacatt tgcatacagt ggctgggaaa ggcagacca 282120  
 cccttaatct ggggtggcac catctaatca gctgccagca cagccaggac attaagcagg 282180  
 tcgaaaaaaa tgtgaaaagg ctgactggc ctgacctccc agcttacatc tttctcccg 282240  
 gctggatgct tctgcactc gaacatcgga ctccaagttc ttcagctttg ggacttggac 282300  
 tggcttctct gctcctcagc ttgcagacaa tctaaagcgg gacctgtga tcatgtgagt 282360  
 taatattctc taataaaact tctattatc tatctatcta tctatctatc tatctatcta 282420  
 tctatctata tctattact tctgtccctc tagagaacct tgactaaaac acatagtaac 282480  
 ccacagtatt atttaccac atatagtaac cattagctta ttttaagccat accacaaacc 282540  
 tatgatgcag ttgttatttt cctcatata aacgagaaat ctgaggatga cagaagttaa 282600  
 gccatttgct tgagagtttt tagctaataa gtgaccagga cagaaattaa cttatctcta 282660  
 taaaatgtta tttttgttaa tattgaatct cagctatagt aaacaggata ttttatgtac 282720  
 agactgggat atatgtgaaa tgattttacc ttattttatat tccaggaaaag aagaaaaatg 282780  
 attacgttct tctcaggttg ttacttcata aacgggtataa ctttttattt tgggatatta 282840  
 tgacaagaat atttgggctt tggctgaagt ggaaggactt gggactacat gtcaaatttt 282900  
 attttgatag taaatattgt gtccaagggt gccatcctt taaagtatgt aaattcagg 282960  
 ttacccctct gtgattatag aagacttata tctgtactt gataatggga aattaactaa 283020  
 atgattttcc atataatgac cattagcttc tcaaaggaac taaaaaagt ttaatgagaa 283080  
 aacacacatt tttattgact acaaaccata atttttgatg gtctcataat caatttaaat 283140  
 atagggcat atttataaat atcatttgca tttgtgttg gatttttatt tgatatatta 283200  
 tttattttaa tttttttatt tcaatagctt ttgggataca ggtgtttgtt ggttacatgc 283260  
 atgaattgta tagtagtgaa gtctgagatt ttaacacacc agtcaactca atactgtaca 283320  
 ttgtatccaa tatgtcattt tttatccctc accaaacctc cctcttctga gtctccaaag 283380  
 tccattatat tactctacat tttgcatacc cagagttccc acttacaagt gacaacatac 283440  
 agtatttaaat ttctcattcc cacttgaat aatggcctcc agctccgtcc 283500  
 aagttgctgc aaaagacatt atgtttatgg ctgattagta ttccatgggt tatctatacc 283560  
 acattttctt ttcttttctt tttttttttt tttttgagac agagtctcac tctgtcacc 283620  
 aggtcggagt gcaactggcac aatctcagct cactgcaacc tctgcccccc gaggtcaaga 283680



gattctcctg	tctcagcctc	ccaagtagct	gggactacag	gcacctgcc	cctcacctgg	283740
ctaatttttg	tatttttagt	agagacagg	tttcaccgtc	ttgccaggc	tggctctgaa	283800
ctoctaacct	cgtgatccac	ctccctccct	cggcctccca	aagtgttgag	attacaggca	283860
taagccaccg	cgcccagcct	ataccacatt	ttctttatct	actcactggt	ggatggggac	283920
ttagaacggt	tocatatttg	caattgtgaa	ttgtgctgca	ataaacatac	acctgtagat	283980
gtcttttttaa	tataatgact	tcttttttatt	tgggtagaca	cccagtagtg	ggattgatgg	284040
attgaatagt	agatccctt	ttatttcttg	actaaatctc	tacactggtt	cccatcgaga	284100
ttgtactaat	gtacattccc	accagcaatg	tttaagcatt	gccttttcac	catattcaca	284160
gaaacggcta	ttgtttggtg	actttttaat	gattgcattt	ctgcaggagt	gagatgggat	284220
ctcattggtg	tcttaatttg	cattttcctg	ataactcata	tgtttgttag	ccatttggat	284280
ttcttttttt	gagaaatgct	tgttcatgtc	atttgcacac	ttttgatggt	ttttttttct	284340
tgctgatttg	tttgagttat	ttgtagattc	tagatattaa	tattttgtct	gatgcacagt	284400
ttgcaaatat	tttctcttat	tctggagatc	gtctatttgt	tctgataact	atttctttgc	284460
tgtgcagcag	tttttttagtt	taatgtgggc	caattaattt	attttgttta	cgtttgcatt	284520
tgcttttgag	gtcttattca	ttaattcttt	gcctaggcct	atgtccagaa	gagttttttc	284580
taggttaact	tctagaaatt	tatagtttca	ggtcttagat	ttcagtcctt	gatccatctt	284640
taatggatgt	atgtataagt	tgagagatcc	agtttctttc	ttctacatgt	ggctagccag	284700
ttttcccgag	accattttatt	aaacagtggg	tactttcccc	aatttatgat	tttgtacgtt	284760
ttgtcaaaat	taagtacagt	gtaagtattt	ggattttattc	ctgggttctc	tatactgttc	284820
ctttgggtgta	tgtgtctact	gtatactggt	accacgttgt	tttgaaact	atagccttgt	284880
agtatagttt	caagtccagt	aacgtgatgg	ctccagattt	gttcttattg	catagaattg	284940
ctttttttcg	ggctcttatt	ttggttccat	atgaatttta	cgtttttaaa	aagtattgta	285000
aaaatgatat	tgggtattttg	atagcaattg	cattcaattt	gtagattggt	ttggtcagta	285060
tgggtcatttt	catgatattc	attcttccaa	cccatgagca	tgggatgtga	ttccatttgt	285120
ttgtttcatc	tacaattttg	ttcagcaata	ttttgtagtt	ctccttgtat	agatctttca	285180
cctccttggt	taactatatt	accaggcatt	ttattttttt	atttttgtat	gtttgcagct	285240
gttgtgaaag	gggttgagtt	cttaatttga	ttctcagctt	ggttgttgtt	ggtgcatagc	285300
aatgctaagt	atttgtgtac	aatgattttg	ttaaattgaga	attcactaaa	tttggttatc	285360
aaatctagga	gtctttcttt	atgatgtaaa	agagaagcaa	aggccaatgg	agtgtcaccc	285420
tctaagccta	cccttacatc	ccagttacgt	tgctaactca	gaaaaatcag	ctgaagggca	285480
cttggaattt	ccctactctg	acccagatag	tcttttgatt	aaaatcatac	ggttattagt	285540
ttttattttt	cagatggaat	aaatgaagac	taataagaatg	tttatgtgga	aactgtgttt	285600
cccaaaaccc	ttctgcagta	gtcttttcac	tgctgaaatc	gttgatatac	taccattata	285660
aaacaaacaca	ctcttttattt	gttcacctaa	cctgcacgac	tttattcctt	agaagatttt	285720
caatattttta	ttttaaagca	tactggaaag	aacagaagca	ttaaagaaaa	aaaataaata	285780
aagggaactcc	ccccacctcc	accaacacac	acaggtatcc	tctttctcaa	ttcatcagtg	285840
aggacgcttg	gtcagtaaaag	gaatttctca	tgatccctcc	tcttttctct	acagggcata	285900
gatttgcaca	cgtgaagcca	ggcttcacat	gactgtattc	cataatagca	cacccaaaaac	285960
actgtcaaaa	tcaatgaaaa	cttcacaata	taaagaatat	agctagaatc	atggtaatga	286020
tcagttgatt	gatataatat	tagataagtg	caggatggta	acttttataa	actaaactttt	286080
caaatggctc	agtgaattgt	ctaacattcc	atggcaaat	tgtatgaaga	ctacctggaa	286140
ctcagctatc	tacaattttg	ttggcagggg	aactagggtt	gtgagtttgt	ttccccctc	286200
tatgtttctaa	tgctcagcaa	actgtgaaat	tttccatgaa	atataatagt	tacgtaaata	286260
taaaaaagca	aaactttttg	gtatgttaat	tattgcacca	acttgtgttg	tttgactact	286320
attaacaata	aactaaaaca	attctacctg	cttgattgaa	gcaaaaatgt	caatgtgtaa	286380
ggatgatggg	tagatcacaa	aatcaaagaa	acagctacat	agtctgactc	agacaggaca	286440
tgtcaaacag	catccctgta	aggaaacttag	tctacagact	cacataggcc	ctgcctcagg	286500
aataaataag	ctgcacgtgt	gttttgcttc	cttgatttct	ttgatgtcag	ggtcagatac	286560
tgggaaagag	actcagggca	aaaaaagagt	atattgattg	acagcttcc	caagacagcc	286620
cacaatagaa	aagacatttc	ttttcaaaga	aactgagtat	aattgtcaaa	aaagaaaata	286680
gatactggtc	atcaaagcaa	gaccttatac	tttgtcaaa	tttactttta	atgtcacctc	286740
ctccaaaaag	ccattttctc	attcccttaa	atcatctca	gtatatctca	gctttcatta	286800
aacaagcatg	ggccatgtat	ttaagaaaa	agaaatttacg	ggaggacaag	acagaccaga	286860
ccttaatttc	agggacata	aattttgcca	aagatgatat	tgccatacta	ctatttctgt	286920
gattctctta	ggatcctgat	gccttggtgc	ttagttatgt	ttttctttaa	tcagaagttt	286980
aaaaagtata	ttatatttta	ttttaagttc	tgtgatacat	gtgcaggaca	tgaaggtttg	287040
ttacataggt	aagcattttg	catggtggtt	tgccacactt	atcaatccat	cacctaggta	287100
ttaaagcccag	catgcattag	ctattttatc	tgatgctctc	cctccccaca	accccgaat	287160
agaccccag	gtgtgttttc	cctccctgtg	tccatatgtt	tttattgttc	tgctccact	287220
tataagtgac	aacatgtagt	atttggtttt	ctgtttttgt	gttagtttgc	tgaggataat	287280
ggcttccagt	tccatccatg	tccctgcaaa	ggacatgata	tcattccttt	ttatgggtgc	287340



atagtattcc	atgatgtata	tgtaccacat	tttctttatc	cagtctatca	ttgatgggta	287400
tctgagttga	ttccatgtat	ttgctatggg	gaatagtgtc	gcaatgaaca	tatgtgtgca	287460
tatcttaaac	aaatttgtaa	gaaaaaaaca	aacaacccca	tcaaaaagta	ggcaaaggac	287520
atgaacagat	acttctcaga	agaagacatt	catgcagcca	acaaacaaga	aaaaaatctc	287580
aatatcactg	gttatttagag	aaaagcaaat	ctaaaccaca	agataccact	tcacaccagt	287640
cagaatggtg	attattaaaa	agtcaagaaa	caacagatgc	tggcgaggct	gtggagaaat	287700
aggaacactt	ttacactgtt	ggtggcaata	tcaattagtt	caaccattgt	ggaaggtagt	287760
ggggcaattc	ctcaaagaca	tagaaccaga	aataccattt	tactcagcaa	tcacattact	287820
gggtatatgt	ccaaaggaat	ataaatcatt	ctgttatcag	gaaattttta	agcttctttg	287880
tgatgtgttt	tattagttaa	aacaataatt	tattagttaa	aatgttgtgt	ctttcctaac	287940
acttaccaca	tgggtgtatc	tgcaatgcac	actaatgaac	gcttcaaaat	tagccttaaa	288000
tgacagactc	cctcacttgc	actctttgct	tatcttgtgt	ctgatggctc	ttctgcata	288060
acaggaaaaat	ctgaattctc	acaattatct	ctaaagcttt	tattaccttc	ccttaaccca	288120
agcccgattt	tattacctgt	atgacactta	gacagtttac	attgtaatcc	cacctcttaa	288180
tggatttaaa	ctcattttcc	actttgggat	ccccatattt	agtcattatt	tgtacttctg	288240
ttattgcaga	cccactccac	cttgtgttgc	agtgtgtgta	tctgtgatcc	cctgtagagt	288300
gtgagttcct	agagtcagag	tatccaggga	tagctctctt	taaaccatcta	atgggattct	288360
atgaacataa	tgaataaaca	tttatgttag	ctaaatgtat	aaagagaaaa	tatatacttg	288420
cccccttatt	aaagttgttg	tttcatatta	ttatatgcca	acaccatact	agtcctattg	288480
acagagagtt	atatgtgtga	tttgtattac	ctttaaagat	attagagcat	aaaataaattg	288540
gaataaactg	ttaggacaaa	ataaagaggt	acaatacaga	cactgaaatg	ggattggcca	288600
attctactta	ctgggacaga	aaagctttct	agaatagatg	aatctcaaag	gataaacgga	288660
ggcataatcc	tgtcagacaa	gaagaaaagt	cgcttccagg	tgaagaagac	agtggtagca	288720
cgaacaaata	tacaaaggat	gaggaagggtg	gagcatccaa	tttaaggatc	aaaataacaa	288780
ataattaaat	taggttattt	agaaggtaag	atcatagata	aacacacaca	tgcatgtgtg	288840
taatggaatt	taaaatcatt	tttttaggtaa	gaaagattcg	ttggaatgtt	ttagagggaa	288900
gaattaagta	attctataca	cattttagaa	agctcattct	gatactctat	aaagaatgtg	288960
tcaagtgtga	aaattagtaa	gatgaagagt	cagtaaatgt	gaatctaaaa	atgtgaatga	289020
gcacaagaga	ttattacata	tatatgtgtg	tgtgtgtgtg	tgtgtgtata	catatatata	289080
tatttttaat	ccaaagcata	tggttaactga	taacaggcgg	gagtattaga	aagggtcatc	289140
caggttgaat	ccaggatttc	atgggagata	gaagggaatga	acagtgtttc	cagacactga	289200
gattaagaat	gtgagaaaag	aagcagttgt	aagtgtgttg	aacttttgac	gtctatacaa	289260
ggagtctact	ggttactaga	aggtttacaa	agagaccatg	gtggatatta	ctagtctttg	289320
tgacattata	gtgtactggt	ataactcttg	gtgctatttg	gggttgctaa	aaatgaaata	289380
aggaagggaa	ggggacagaa	gataaaaatg	gagggaggat	aaaggggaga	acatgagagt	289440
ggcaagttaa	aaagtagggc	ctagagcaga	agacagaagg	aaagcgaggc	atacaaggta	289500
agggaaaggt	aagaaggagg	cgagaagaga	acctgtgtga	tttccaacat	ctaagtgtgg	289560
acaggagaga	ttgtctgtaa	aagggagaag	acatggtcac	gaggaaaggg	ggtgttagca	289620
aattttcatc	atggaagcct	aaggtaaaaat	tttcagagaa	ggatgtttta	tgtaagtgtg	289680
atctaaaatg	gccaaagatc	atcatgatga	taagtgaaca	ttcagataaa	gtttcagaac	289740
agataaaact	ctgtaatggt	gacagcagcc	atcggagtgg	ttacatggta	tggcctgggg	289800
agggatgtgg	ggaatcctgg	ggtgttgaa	atggcctatg	tctcaaagtg	gtgaccaatc	289860
acattaccgc	acttcattgt	ctttgtctata	agcaggtcac	ccctcgattc	aaaatgtgtt	289920
aaaatgtaat	acgaaattaa	tataaacacct	aagagggtta	aaaaatattag	tggctttctc	289980
ggattacctg	ggtaccatca	gctgaaggat	tcaaaccaaa	ctctgttaga	ctcagatgcc	290040
ttagctgctg	accgctgtac	tactctgctg	tcaaattccaa	tccttgccaa	gcctctccat	290100
tggcatgtta	aatggactta	gtctctgttg	tcctttttag	tcttattttg	aaggcaccta	290160
cctattctac	caactgataa	attgaattgt	tattactggt	ttacagtaag	atttggtag	290220
aagctttggt	tagaccataa	ttttttgtca	gatctgtgtg	tctgaactct	ccattaagga	290280
actgtttaga	tattacatat	acttttgctg	aaattcagag	gtactcataa	atttgacgac	290340
ttttaaaaaa	atgtactatc	agcagtggtg	aatgataact	aaaatggtat	aatttatctt	290400
aacagtatct	tgagcttaac	aaagcttcag	ataaaaacaa	gattgcattg	aatatcatgc	290460
aataaaaagga	aatacttaaa	gcaaaggat	ttatatgttc	ttatgtgtat	gttcttacag	290520
ggatggaaaa	gtctattttc	ataaaaatcta	attttttaat	ttacaatata	gagccctttg	290580
cttcttttcc	tgttttaggt	aagatctttt	cttctcttga	aagattatgt	atattttta	290640
atgattaatc	ttattttacc	tttacagaag	tactgtgaat	tagcagggca	gatattatga	290700
catctcctta	acatatgaat	taattgtggg	ctgccaaata	gtacatgatc	ttttactgga	290760
acgtagatcc	aggtggctca	aatagatatt	taagggcaga	gcacctaaac	tgaccagcca	290820
atgtaagcag	ttcatcttcc	gagcagattt	atttttcact	gtaaccaaat	gtattttcat	290880
gctcctggaa	aaagtgttct	cccgtgagct	ttgttagtgt	gttttcacca	aggaactgct	290940
atctggggtg	cccacactgt	gtagagtttt	tgtcctgaaa	agctaaagtc	catatatatga	291000

agaaaaccta	ccacataaag	atgtttccta	gaatacccgt	catgatcaaa	ggtttgcagc	291060
tgtagtggaa	ctaactaatg	gatgcttcat	tttagtaagg	gcctcttttc	ttccattgat	291120
catagttgaa	tattgtgggt	tcagtttaag	atgagtcctat	gtgtgacatg	gattttaaata	291180
cccaaatctt	agtggtttga	taaaaattat	aagaaatact	gcaattatct	caaactaaaa	291240
cttgtcaaga	tttgtgtaat	atcatagaat	gtggacaaaag	tgattaaaaa	atatattcta	291300
ttgactctgt	gacataatct	tagcagatac	atctttatat	tggccaacct	ttctttgaat	291360
taatcttgac	ttacgagttt	agaatagtg	catgtctgga	gagaaaaaga	cttcatgtga	291420
aatataaaca	ctctatgtac	tatttttagta	ttaatttttag	aaagatattt	ggcacacctt	291480
gaaacgtgat	ctaataatcac	tttcaattct	aagtctattg	catcatcaag	cagaattata	291540
taaaactattt	ttaaaagttt	atcaactgga	ttttatgtta	ttagcaatac	atgagatgtc	291600
atggcttttaa	aaggatggct	atttttattat	gtagtatcct	tcaaataatta	tttttagtagt	291660
tttttgtact	aaccataggg	caattaatga	ttcagaaata	tattttaattt	ctattttata	291720
aatctttatat	tcattggcaat	atactacccc	tcactgattt	ttgcttggct	ttattaaact	291780
gagtttctgt	tttatctctg	caaaaataaa	taaaaaataa	aaatatgctc	aaatattaag	291840
caccccataa	atattaggtg	gataaataaa	tataacccaa	ctgctcttgc	aaaagtgaag	291900
aagtttttaaa	tatatagatt	tcattccatt	ccttttaatt	tttcaaaata	ttagatgata	291960
caaaaatcat	gtgcacaatt	cagactcata	gctattttta	gaaactataa	tttttgattc	292020
aaaagaaatc	tgcaatcttg	caaaaatgag	actcagttta	acagtgtatt	tccaaagaaa	292080
ttgtagctgt	tattcttgcc	tttcttctta	gacaagagac	ttgtgatgg	atctcttaat	292140
attccatagc	taacaagaaa	aagagcagaa	aaagccatgg	gaactataaa	tataggcagt	292200
gatatttggtg	tattgtggtt	gaaaacatta	ctctggaagt	agaccacaa	aggttctcgt	292260
tctagctctc	catctaactt	ccatctctag	agatgtgccc	ttgacaaaat	gattttttcc	292320
tctggacttt	aatttcttta	tctggaccag	gtaagctcaa	agatctcagt	ttagtatttg	292380
atttttccag	attaaaatct	aagatgtggt	atcagataca	cttttctatt	tttcttttat	292440
ttgtcagaca	aacaaggtga	aaatgtgttt	aaaaactagg	tttatccatt	tttttctcgt	292500
acagtgttcc	taattcaaca	ctacattttt	ctatatactt	ttagtagaaa	gtgaatcata	292560
agttacttca	gtaggatata	tctgttgttt	tgctcttgta	ctcttccttt	actattacta	292620
ttatttcata	acataaataa	tgtaagaga	tctgggtcta	aatgtaggat	ttaatttttc	292680
aagttttttac	aaattctatg	aattccctac	agtttctttg	gcogtcacta	tttgtctggt	292740
tgtttataga	agccgagaag	cagaacccct	atggctgaag	actggctcctg	atagagattc	292800
ttcttctaatt	cttctcaaac	caggaagggtg	ttcaggtcac	tctgtaaatg	attcaaagac	292860
agatactatg	taaagttaca	tttttctctt	aaaatccagt	ctgtaagcaa	attttatgcc	292920
ttatttttaaa	taacaagata	tgtagctttg	gcaacatgtc	atcttatttg	gataggctat	292980
gccaacagga	ccaacaccaga	gcaagccaaa	aaagatcttt	gaatgacaga	cagcttaaga	293040
agaacagaaat	ctgttagaga	aaataagaga	aaaggagaaa	tgtaactgagc	agacccaaaa	293100
aataaaataa	aatagaataa	aaggagcttc	cacacaaggc	tgttgggttc	attcaccaca	293160
taccaaggga	caactaccaa	gtaaaattta	taaaataaat	acttgctgga	gaaactgaat	293220
atccacctgt	aaaataatga	atcagaacac	ctacttcact	tcattgccata	tacaaaaact	293280
aaocccaaaat	agatcaaaga	tctaaattta	ataagtataa	atataaaact	cctagaaaga	293340
aacataggtg	aaaatatgca	aaagttttta	ttagggaatg	agttcttata	tatgatgcca	293400
aaagcacaag	caaccaaagg	aaaatcgata	aattagattt	tatcaaaatt	aaaaactttt	293460
atgcattataa	gaatactatc	aagaaaatga	aaagcaagta	acagaatagg	agaaaaatatt	293520
tgcaaacata	tgctcgacaa	gtattaagta	tctagaatgc	ttcaagaact	cttacaaatc	293580
agcagtaaaa	acacaattca	attaaaagat	gagcaaagtt	caatattaag	ctattcccca	293640
gagaagatac	acaaatggcc	aataagcagt	gaaaagatac	tagcattagt	cattagagaa	293700
tgcaaatcaa	acctacaatg	aaatagcact	ttgcacctac	taggtggcta	taattaaaaa	293760
ataagcaaac	agaaacaaaa	caactattgg	tgtggatgta	gagtaatcgg	aatcctcata	293820
cattgctagt	aggaatgaaa	catggtacag	ccactatgaa	aaaagaaatg	ttttcagttc	293880
ctcaacaagt	taaacataga	attaccatat	gaccagcaa	ttccgcagct	aggtaaatgc	293940
ctgagagaag	tgtgttagtc	cattttccaa	ctgctgataa	acacatcccc	ggggtgggt	294000
aatttataac	ataaaagagg	tttaattggac	tcccagttcc	acgtggctga	ggagggcctc	294060
actocacaatc	acagcagaag	gtgaaagaca	tgtcttacct	ggcagccagc	aagagaggat	294120
gagagccaag	tgaaaggggc	tttgccctta	taaaaccatt	gagatctcat	cagacttatt	294180
cactaccatg	agaacagtat	ggagaaaacc	accccatga	ttcaattatc	tcccaacctg	294240
tcctttccac	atgtgggaag	tatggttaacc	acaattcaag	atgaaatttg	ggtggggaca	294300
caacaaagcc	atatctgaag	taagacaaat	tatctcacia	aactatacac	caatattcct	294360
agcagcatat	tcatttatagg	taaaagaaga	aaaaaggcac	atgtccatta	tattcatata	294420
atagaataatt	attccatcat	aaaaagggaat	gcggtagtgt	tacatactac	atgatgaacc	294480
ttgaaaatat	tatgctaagt	gaaacaagtc	tgacaogaaa	gaccacataa	tgatgatctc	294540
catacagaag	aaatgttcag	aacaagtaaa	ggcacagagg	cataaaatag	attagtgggt	294600
gccagtggct	ggcgggagat	ttccattaac	agctaacaatg	tacagggttt	ctttttgggt	294660

tatcagaaaa	ttctggattt	catatttatt	gttgagagc	atagtgaata	tactaaaaa	294720
aacaaacaaa	acactaacat	atactcttgt	aactgggtgaa	ttttgcttat	gtgaaaaata	294780
tgctcagtaaa	aatgctaaag	taaaacaacc	tgaagagttg	actatatata	tttgtagata	294840
tacacaagta	tgcathtagc	ttttataaac	aatgtcttgt	taaattatta	ctagcaattt	294900
tatagtcagt	taagattgaa	tacatgttta	attcctgcct	ttattgaata	ttcattctgc	294960
aactacttgt	gctgggcact	gtcctaggtt	caaataattt	aacaataatt	aagagagaca	295020
tgcttcttct	agaaaagac	aatgatttga	aaacattttt	atttcaggac	acttttatac	295080
tcttgaaaat	tattgaggac	cccaaagaac	tattttttat	ataggataca	tccatcaata	295140
ttcactttat	tagatattaa	actgagaaaa	aatttttaagc	atgtgttgaa	ataataaact	295200
attttaatat	aaataatgca	tttattaaaa	tattttctat	tagaaacatt	ttagagaaga	295260
gtaaaattat	ttcatatttt	tgcaaatatt	ttgacatctg	actgaagaga	agattttcag	295320
atttttacat	ctgctccagc	atttagtcta	ttgcaggctc	ttgttttggg	taaaatgtat	295380
gaaggaaatt	ttgccttgca	cggatatgga	aaagagagga	gtatttcaat	taccttttca	295440
gaaaattgtg	aatatatatg	tatatatatg	tgtgtgtata	tatatgtgta	cagatagata	295500
gatagatgat	agatagataa	ctacatcagc	actaaacaag	tggaatttc	ttacatgttt	295560
gttgcaatgt	gtaatccgga	actgtatcag	tgaaattctt	ataacctgtt	acattaaaa	295620
tgattggtcg	ttctttcact	ttgaagagat	cttttaatca	aaagtgattt	tgcaaatatt	295680
tgcatgtgtc	atttgaaaaa	ggtttgttca	tgaagctttg	cagattttcc	aattgttgat	295740
acattttatg	atacaaataa	cacttggttaa	taccaccact	gttctcaaaa	ggagttttaa	295800
gtgtggaagc	cgaagctgac	caagctaatt	gtggcacaca	cacgttttcc	gaaattctgc	295860
ttttcacttg	aaagctagaa	ttttataatg	aataacaagt	actgtcagat	gtttccctta	295920
gggtcacttt	gttcatgttt	gaaaaagaaa	agtctgccaa	atgcctactt	atgaattacc	295980
atggtttgtc	agtcattctc	tcaattagaa	gcgaagagct	acaaaaagt	gggcagcgaa	296040
ctctcactcc	aacgaccgca	tgagcgcccc	ctgcccgtcaa	ctgtcacgct	gccgtgcagc	296100
agaagtgtcg	gatttcatct	gaggatgtga	aaaaggcatg	tactcaggat	ggaaattcca	296160
cataatgaac	acttttcgct	gcttcatcaa	ggacgtgctt	aactaaaagc	agcatcggtt	296220
ttggttgttt	gtttacactg	gggtgtgtagc	agtgaagact	tcagtggcac	ctagtgcggt	296280
agctcctcga	ttcttcatct	gcactttcgc	cattgatgcg	ccactagttt	tgctccccct	296340
atgcaatgtt	agcacagtga	aaaagacaaa	tacatttcgt	tttgttttgt	tttttttgag	296400
atggagtctc	tctctggcgc	caaggctgga	gtgcagtggc	gtgatctcgg	ctcaccacaa	296460
gctccgcctc	ccgggttccc	gccattctcc	tgtttcagcc	tcctctccga	gtagtgtagc	296520
tgggactaca	ggcgcccgcc	gccacgcccc	gctaattttt	gtatttttag	tagagacagg	296580
gtttcaccat	gttgacacag	gtgtctgca	tctctttacc	tcgtgatccg	cccaattcgg	296640
cctctcaaa	tggtgggatt	acaggcgtaa	acgtttctta	atattatggc	aataaatata	296700
gttttggtct	cagagtcccc	ctaaaagtgt	attgggacac	ccagagacct	gcctgagcac	296760
cataactgga	taactgctgc	tctagattct	aaaactttct	ctgagttgct	ttatcaattg	296820
tcctttctct	gttttcttac	acatgggaga	ttcttctca	ttgtctctga	aaatcctcat	296880
gtttcttatg	gttattttat	atggcttcat	gatgcacaca	cacatcotta	tatatatttt	296940
cttccgaatc	ttctctaaca	tgccctcttc	ttgcctatag	atcatgacag	tattctacgt	297000
acatattttg	aggagttttt	attatgttca	ctctgggaga	gctttcgctc	ctcagggtac	297060
aggagacttg	tgaaccccag	agtcagtggc	agtgcctgca	ctccctagct	tgctgctcac	297120
atcggtagct	gaaatgcctt	agtcttgcct	cagttatttc	ctttaaatag	atgctaggtc	297180
gaataggtgt	tgtggcctct	tagtaactga	gtgatgggag	aggtcatttc	tttgcttata	297240
agaacacact	gcttttctaag	tgagaagagc	ccagcttcc	gacagagcac	cattagaaga	297300
accctccact	gtccttggtta	ggttatagtg	cttatcaatt	tctctccttt	acagggccca	297360
atocaatgag	tgacctgttt	tggaagcaa	acatgtttaa	agaaacccta	cttgaaagtt	297420
aaagactaca	attgcatttt	agaatataga	ctagtgtctac	caattgtgat	tgctcctcctg	297480
aacgctatgc	ttttgtgaca	gttaatggaa	tttaaaaagg	aaaagtaaac	gtatctataa	297540
cctaactttt	gtaatggaaa	ggattggatt	ggaatggaa	cctaatacat	tatttctctc	297600
agaattaagt	ccattgttct	taacctgtta	ttcttcagtg	acctcgcagg	actgcaagct	297660
acctaaaact	gaatgtaaga	ttttgtgagt	tttttttttt	tttccctgca	tgatatatag	297720
acatatatac	actggggaaa	gaatatcaaa	aggtacattt	gttttattca	ttcaaatatt	297780
aaatatccaa	atatttctta	tatacacagg	tataggtatt	tagttattgg	atggtaaaat	297840
atatatattt	ttatatgtat	gttttgggtca	tttcaacttg	taaaatagag	catcagtggtg	297900
cactatggac	agaagcagta	ccatagtttt	ttcctcttat	atcaactttt	cttctcattc	297960
cctctttttt	cttctgatcc	atctccttcc	cctccctcaa	ttccagtttg	tttttctctt	298020
ttaaagttgg	ctatggacag	caaagacaa	tgaaatgtaa	tttaaaatc	gcacatttaa	298080
catattctat	attataaaatc	atgctgctat	aaagacacct	gcacacttat	gtttattgctg	298140
gcactattca	caatagcaaa	gacttggaac	caatccaaat	gtccaacaat	gatagactgg	298200
attaagaaaa	tgtgtcacat	atacaccatg	gaatactatg	cagccataaa	aaatgatgag	298260
ttcatgtcct	ttgtagggac	atggatgaaa	ttggaagtca	tcattctcag	taaactatcg	298320

caaggagaaa	aaaacaaaaca	cogcatgttc	tcactcatag	gtgggaattg	aacaatgaga	298380
acacatggac	acaggaaggg	gaacatcaca	ctctggggac	tggtgtgggg	tagggggagg	298440
ggggagggat	agcattaggc	gatataccta	atgctaaatg	acgagttaat	gggtgcagca	298500
caccagcatg	gcacatgtat	acatatgtaa	ctaacctgca	cattgtgcac	atgtacccta	298560
aaactttaag	tataataata	aaaaaaaaatt	cttcattgac	ttacatgcaa	tgaataattt	298620
ataattagcc	tatgttacta	ttcattttcc	ttttattttt	ttcttgcaaa	attatgtttag	298680
aactgtgact	gaaatgtatg	agttttctcaa	tagatgctat	tgattgagtt	attttgaagc	298740
acaaaaataa	attgaattgg	ttaatatata	ctttgttaggt	caaaactggc	cctgggaatt	298800
attttttaga	gctattccac	attattcaac	ttcgtaagct	tactgttggg	tatttatata	298860
tttttaattt	ttgaattata	atagaagcag	tttaggactt	aaatataaaa	atatgtaaaa	298920
tgtaaatata	ctcatttttc	attttaaaaca	catgctttca	tgtcataacg	gatcccaaaa	298980
taccaaagat	attaggggaa	aaaataaaact	gttggtgtct	taaaatacat	tcctttgttg	299040
aatctatcct	gctaaccctt	taggggtattg	acagcctatt	atttttcata	tgtgccatct	299100
ttatctagaa	aactgagaga	acaactaatt	atatcctttt	gcaaatagag	gcaacagaat	299160
ggcttttggt	caacacaatt	tttaagtctg	ttgaacagga	ctggtgaatg	gcttaaaaaat	299220
ttgccatggg	gtgtgtgtgt	gcacatatga	gaatatgtat	gttaatatct	agccacctta	299280
gtgtcatgga	aagcgtggcc	tggggatttg	attatacatt	gtgctttgtc	ataaatttta	299340
tggtttcttt	gcaacattat	tggaggaggt	atgcctctat	gtctttcatt	tagcttgtat	299400
atttttggta	tgatagacat	ggtaattact	gctcttaaaag	tttgtgcaaa	tgccactggg	299460
aaaggggata	ggagtaaaga	attaaagtga	aggctcatgaa	ataaacctat	taacaaaagc	299520
agagtgattt	cttgaataat	cccttcatct	ctccccatgt	cataatcctt	tctcataaat	299580
aattatacaa	gcaagcctag	acacacacta	gtgttgaaaga	attgtgctgc	aaaatgcatt	299640
tagctgcatt	aaataccaag	aggaaacac	agttccaaaa	cctccttaaa	gtggaattct	299700
cagttgtttc	cagagaagag	gcaagttata	gtacatataa	tctattatga	acgtatctga	299760
gtagaacacc	atgcaataag	gcgctattga	attgggaggg	aggtatcaga	gctaaagagc	299820
acaaagacgg	gccatcgaga	tattagagga	gatttctctt	catatgtaat	tctgctttac	299880
atgtttctct	aatcttattt	cctttctccc	taccgcctct	ttttttttta	gccacaaatc	299940
attgogtttg	tacagattac	aggaaaacac	tcaaatattg	ctcgaagaag	cataatcttt	300000
cttctcgagt	tggtaataaa	tgagatctac	ttgtaaaaaat	tagtgtcatt	gtctacaaca	300060
ctttcagatg	actctcatgc	caccaggtca	atattcctgc	tagggatgag	ttagttctta	300120
cacttggttg	atcaccaaga	atgtgggact	cataagggca	atggatgatg	caggcgcata	300180
gtgtcctttg	atgttactgg	aggtgtaaaa	tgaaggtttt	ttttcattct	aatcaaaagt	300240
ttaatgggat	agtaagtatt	tattatcttt	atcctggaag	ggattttatt	gcttttgaca	300300
tttttaaaaa	tgtaaaattt	taattttaat	ttaatgtgtt	aaaaaattt	ttaaactttt	300360
atatrgtttt	acaattcagt	ggaattctct	cttctttata	ttattgtact	tgcatataa	300420
atatttttaga	tgttttcaat	catgagttat	tgtcactatg	ttttattaga	aatatccttc	300480
tcactatgag	ataaaggggac	gaaatgagtg	ctaattgtaat	tctacagggg	tcagatagca	300540
tttccatggg	tacatttgatt	cactagcagc	atgtcctgaa	taactcctgc	tgtattgcaa	300600
acccttgctg	ggcagggctc	agtatatatt	taatgtttct	gcatatagct	ccaatttagc	300660
tccaattttc	ttcctacatt	gtagtactag	atcaaaaaat	acttattggg	tgaacctata	300720
atgtatgcta	cttaggcagc	tctggcatga	ggcatgacat	tatattttaa	ataaagaaac	300780
acattccagc	attccaaaaat	tctcaaagca	taacctgtg	taaaaccata	gttaaagttt	300840
caaaatttga	aacacgaaaa	ctgacttttt	aaaatgttaa	ccatacaatg	actaatgcag	300900
tgtgctcttt	gcacgtgcat	aaacaataag	aaacattatt	aaaataaaaat	atttatattt	300960
taaagttaac	atacactatt	taaaatactg	agcaaacatt	tatataagta	ggggattttat	301020
ataaaaatctg	aatgatgaaa	gtacttaaaa	agtactaata	gtaactaatt	tattgaaatc	301080
ttgtggcagt	cttcatatga	ataatttaac	caccagaata	gttacgggaag	gtaggtacaa	301140
atgaataacc	cattttgtata	ggagttaaact	gatgcctaga	aagagaaaagt	cactaattca	301200
agctcataca	acgtacaagt	agaaaagtga	ggatttttagc	acaggtaatc	ttgatacaga	301260
atgatttcca	ttatttttaa	ataagagtaa	tacatgtatg	tatatgaatt	caaataatac	301320
agaatgataa	aaatggcaga	atacggagcc	cccttttcca	cccctcgagt	tccttctccc	301380
caaccaatat	ttatattaat	agcttatttt	atztatctta	aaaatattat	attttacact	301440
cattattttct	atctatatct	ctatatagat	atatatcttc	atcattataa	aataagatat	301500
acagtataaa	acctttctct	taatatgaat	ggatgaaactg	atctaccata	tattattagt	301560
agctatattg	aattccgatg	tatggatggg	cctaattatc	ttttttatta	ataaaaattt	301620
agtttgttta	tatcattaat	attctcatat	tttgctattt	gcaaaaatag	ctggaataaa	301680
caagagttaa	ttttatttct	atttactgta	aatttgttca	ttattactga	agtaaaagga	301740
ttgagatgac	ataatgccta	tcagtgcctt	taaaatattt	caataacaac	aaaaaagaaa	301800
cctctttcaa	gtgtgcaaca	tcctgataat	taacttatct	gggtaatgag	tgtatgaaag	301860
ttcattataa	atgatatttc	tatgtgtgtg	tgtatttaga	atttttacaa	tagttttgtc	301920
aaaagttatt	ttttaatgaa	gcactctgatt	tttagggagg	gatctaagct	ttccacacat	301980

gcttctattc	tccatagtag	tgtaagaatt	tcaaatgtgg	acacgttaaa	gtattttctag	302040
aggaggatta	agtgatatga	atgaaactgt	tggggacctt	acatttttatg	tgtaatgaga	302100
gctgaaagg	ggttctacaa	agaagtgtga	taatagagga	cagatttttg	gaatgaggtg	302160
atgtcattcc	aataaggact	atattttatct	agggatattt	tgaagaaac	tggtgatgag	302220
gagaatgaag	agaaatagtc	attagtgc	acatgtcaag	aaagaaaaca	tattacttct	302280
ttggagatat	tttctctctc	tctctctctc	cccctgcctt	cccctgtccc	tctctccata	302340
aggctctatt	tatctatoca	tcctctctgc	atatgtcttt	atccatactt	atatctccat	302400
actcatatct	ataccgatat	ctatattttat	atctgtaagc	tgatcttagc	ctactcatca	302460
tggttaacata	tggcagattt	gtgtatacct	gttatcagtt	cccaaactctt	tacacacact	302520
tctctgtcat	catgatttaag	gggagtaatg	tgctcttacg	tttttgagaa	ttgggcccaa	302580
tggggacagc	tatttttctc	tggtataaga	ttccatagat	tggtggcaaca	atggcaccat	302640
gatgacttct	aaactcttga	tgggagactt	gcttctgtga	cttaattggat	tattagccct	302700
gcaatgggaa	ctacacaccc	aaaggccaag	aggaaacatgc	aaaggcaaac	agagagacaa	302760
gagagcccat	caatcacgag	aagatggcag	gaagtgtgtc	cgttcccatt	gcagttattc	302820
ggacagctta	agattgaaat	ataatagaaa	gttatatgac	gtactctaca	gacaagtta	302880
tattttatta	ttattattgt	tttgtaatgt	agaaaacatc	taccctctaa	aatctgtgga	302940
taatagggtta	ctagattttg	taactaaata	agataataga	caaaatataa	tagatataga	303000
aaaagtcaga	aatgatcctc	atccattttta	gaaacccaat	caattttcaat	ttaatccaaa	303060
tctcttccct	cttcaacaag	gcactacggt	atcctcatcc	ttccagcttt	ccttcccaca	303120
gctttgtaca	ttcagggtta	gtcttcatgt	cccatacttg	ttcctatttg	ttttaaattg	303180
agtcttaata	aattcttctc	attgggtggg	aataagatat	attcaataag	gtgaaatgtc	303240
taactttaat	tactgaacaa	tgaagcatgg	tagagggtta	tactttaata	atataattta	303300
ttaattatac	atcatgggtc	aagggtctcta	ctttaattaa	tattttccac	tagcaaaaaa	303360
taactatcgg	tgtacaaagt	cccacagatt	agacaccaga	acottaggca	ggtagtaagc	303420
cccagtgagt	tcagtggtta	aagtgggctt	taagtaccta	agggttatct	ttattcatgt	303480
gaaagttttc	tggttagaga	gttaacagct	gagtgcattt	ttagtaaaat	actgcaggaa	303540
atggctgata	aaagtatcaa	tggtgtatat	ctgcttaaa	gagatgcata	tctgttcttt	303600
ctcaaacac	tggttaagctg	tcacccatag	acttattgtg	gaaacaatta	tctcacagta	303660
gtttttcttc	cagacatttt	tctgcccaga	caaaagaatc	tatggaagta	gtcattttctc	303720
tagtttgttt	tagggaggat	tggtgtgggt	ctcctatgag	aaatcacttg	cacattatgt	303780
tcagggtatac	aaacacgtga	atgtatttgc	tatacagagt	ccattttctta	cacacagcaa	303840
ggtatagttt	caaattgctg	aagccaaata	aaccaaatgg	gaagaaattg	atggtaaact	303900
ttgtcagaat	gttgggtctc	tctcatattt	atgaatctgt	caaaggatag	actttttgtg	303960
aactaataag	tttttatatg	ttctaattat	gcaaggaaat	ctatcctgtc	ttagcaatca	304020
agtggttaaga	gagaaacata	tggttagaga	tcagggaaga	atatgttaag	tgaataggac	304080
actgaaaaag	tctgctgtga	atacagtggt	gggcagccat	tgtagagatta	agtagtaaat	304140
gcccagtcga	acagaaaaaa	aaatggaatt	gtaatataaa	gtcacaaata	taatttggtg	304200
attgacttac	acggattaca	aatttggtat	attgagcttt	ccgggcttag	tgaataacat	304260
cactctttgc	atatttttgc	tacatcatct	ggccacaaaa	ttttaaagac	tattttatgaa	304320
agtcagggtta	atattttaaag	tcacaaagag	tcttagaaaat	tttgaacatc	acactgactt	304380
cgtacctgtt	gtactatggc	atatgtcttt	agtcaactaa	ttatttttct	cgtaaacaaa	304440
aaaaaaggaa	ataaaaaatc	ccaaaaaaca	aataacccaa	taaaaataat	gagtaaaata	304500
tcagggttga	tattctcaca	aagaagacat	acaaatgaag	agcaagtata	tgaaaaaatg	304560
ttcaacatca	caaatcaaaa	atcaaaacca	cagttagata	acagctcacc	ccagttagaa	304620
tggaattctc	ccaaaagaca	aaagataacg	tggttgagag	gacgtggaga	aaagagaatc	304680
attgaacatt	gttgggtggg	atgtgaatta	gtatagcctt	tacgaaaaac	agtatggaga	304740
ttcctcatga	aaacaaaaat	agaagtatca	tttgatocaa	cagtcctact	actgggtaga	304800
tatccaaaga	aaatgaaatc	ggtatgtcaa	agaagtgtct	gcgcttccat	gtttatgaca	304860
gcattattca	caatagccaa	ggtatgaaat	caacctaatg	ggacatcaat	ggaaaaatgg	304920
ataaagaaaa	tggtgtatgt	atacacaatg	caatactatt	caaccataaa	aagaatgaaa	304980
tcattgagaca	acaccaatga	acctggaaag	cattatgtta	agtaaaagaa	accaaacata	305040
ggaaaacaaa	caccgcatgt	ttcactcatg	tggaattcaa	atacattggg	cacatgaagg	305100
tagagtacaa	tggtagctgt	cagaggctgg	gtgtttaggg	gagaaatggg	atggagagat	305160
tgcggtcaga	gcatacattt	acatttgata	gaagaaatga	atttcaagag	atctattata	305220
tagcaagggtg	actctaatta	atgaaatat	gattctattc	ttgaaaaatg	tagaatgaa	305280
gtaatatgcc	atcaccacag	aaattgtaac	agtgcagggt	aatgcatttg	tgaattagct	305340
agatttaaca	agtcaacaat	gaatatacac	tttataacat	catgttggtc	atgataaaaa	305400
acatacaatg	ttatctgtca	attaaaaaaa	tatctaggaa	gaatctgtac	agaagattgt	305460
gggtattttat	aaatgcata	tttactttgt	aagtcaccaa	atttgtgaat	aacattattg	305520
gattttatttt	gottaataga	atagaataaa	taatagaata	aattataatt	aaatatttaa	305580
agtgatccct	ttagaccttg	atgattttat	aaatataggt	gagatgctag	tttgcttacc	305640

ctaaatatct	attttctgcc	ccagaaacaa	tcctttaata	gaacacttta	cctttcotta	305700
acttccatca	ccaaaattgg	caattaagaa	tattttcaca	cagaaaaaac	cactattgat	305760
atgtgtgtat	ttttagcaag	aaagccagtt	ttttgagctg	tatgatggag	ccttttcaact	305820
cattagtgat	gcaactgact	tctgcacacc	atcctacaga	agaagaaact	catgttcgta	305880
tataagcaaa	atgacatatc	attacacata	ataactcatt	gtggtgggtt	ctaagcaagt	305940
gaagcagtac	tggctaaate	catcagtgtg	gtattggctc	tctattctac	ctgttcttaa	306000
aaccagtaag	gtagagttcc	cgctccttga	tataagaata	aggattccac	cttctcctgc	306060
ccctcccaac	tagaattctc	aaatgtgttt	gtcgtgggtt	catccaatat	tgcccaagga	306120
aagtgtgttt	tattagcttg	ggattccatc	tgtgaagact	ggcatttcta	tcttactaac	306180
agatacagaa	catatgaaga	atccctgagt	tcctacatag	ctgtcaagaa	gtaaatgaac	306240
tgatttgact	gttattccct	tcactcaactg	gaaaaatagc	tcttcccaga	atcaagtatt	306300
gttgactctg	ttccaatcct	cagagcaatt	cttgccaaac	ttcttttgag	aggtgctgcc	306360
ttgagaaaagc	atgacctatg	aaaatgagtt	gaagctccaa	accagcccat	accacgtgat	306420
ttgcatagta	caagcctcac	tgatgatcac	tatccaccat	atcttagaga	tgggatttag	306480
aagttactca	tctctggaga	ttgtatccca	gggtctagca	caagtctcta	cacaagtggg	306540
actcaaaagt	agcacttgca	ttagtattag	ctgagaattt	tttaaaaaaa	aatcggagaa	306600
cgaagcaatg	aagtaaatgt	agaaaattcc	tcttcatgtg	aaattcctgc	taacatgttg	306660
tacagcatgt	taggatgggg	accacacctc	ctttttactc	accattata	gcccacgat	306720
ggaacacagt	gcctgacaat	aagtatgtgt	tgagtaaatg	aaccgctgtg	atcctccac	306780
tgcataccg	caagtacctt	tagaccccaa	agagcataat	acagatatgg	tttggtatg	306840
tcctcaccga	aatctcatct	tgaattgtaa	tccccagttt	tcttactggg	ttcagaaaga	306900
gaagtatttt	tccccacatg	tggagaaggg	gacctaatgg	gtaactgaat	aatgagggca	306960
gtttccccc	tgccgttctc	gtgatagtga	gttcttacia	gatctgatgg	ttttgtttgt	307020
tatttgtag	ttgttttga	gacacagtct	tgctctgtca	cccaggttgg	agggcagtg	307080
tgtgatctcg	gctcactgca	acctctgcct	cccgagtcca	agcgattctc	ctgtctcagc	307140
ctcccagtag	gctgagacta	caggtgtgtg	ccaccacgcc	cagctaattt	tatgtattct	307200
taatagagat	cgggtttcac	cgtgttagcc	aaggcgtgtg	cgatctcctg	tcctcgtgat	307260
ccgcccacct	cagcctctca	aagtgtgtg	atgacaggtg	tgagccacca	tgcccggtg	307320
atctgatctg	atggttttat	aagcatctgt	catttccctt	gcttgcaatt	ctctctcctg	307380
tcgctgtgtt	aagaaggttc	ttgcttcccc	tttatctttt	gccatgatta	taagtttctt	307440
gtggcctccc	cagccacgtg	gaactgtgag	tcaattaaac	ctcttttctt	tataaaataa	307500
ccagtcttga	gggtatttct	ttatagcagt	gtgaaaacgg	actaatagac	aaagaaagcc	307560
tcaaggaag	aaagtgacat	tgtagaatat	caaaaatatc	cagtaattct	ttttcatagg	307620
aaacgtggct	tgttcattgt	tgttcattga	gaaccacctc	tattggcttt	gtatcttgct	307680
tcaagtgtgt	ctagaaattc	caactattat	tgaggcattt	gcgtattaaa	ttacttttta	307740
gaaatcttta	cagtgaaggt	gaatctggta	ataatacgta	acaacattat	gctaaccact	307800
ttaatgttct	agatgtattt	gataagaacc	ttattctcag	tattctgagt	aaactcgggg	307860
cctggaaaac	atttagcctg	tacaaagatg	tatatgtgtt	atatatttct	taattgaaag	307920
acgatataata	ttttgaaaaa	aacattttagt	ctcctaaaaa	tataaaaaaa	tcattgttta	307980
gagaatgcaa	ctaacttcaa	aattttaaat	aaataaccag	tgcttatagt	gtagtgtgat	308040
tatccattca	taaaattgat	tgcaatatgt	tagtgctggc	agcttagtat	aacacatttt	308100
ccccgaacc	acaaaaatgt	gtttatttat	ttattcattt	taaaactact	ttatagtaat	308160
atacagatta	agtaaacatt	ttagtaatga	gcttacagat	tttatacgat	catgcgggtt	308220
tggttttctt	actgggttca	gaaagagaag	tatttttcag	tggggattgc	ctagctgcaa	308280
ccttgacatt	ttcactgtct	tgctgcaggg	agttcaggag	catgtataag	actattctat	308340
ttaatttagt	tttacaatct	cattgtatct	ggcaatttaa	gacatagaaa	agtaactttt	308400
ccttaattaa	aataaacttt	tggtatttgg	atagcaggat	aatgaatacc	cagacaggca	308460
gtatgtgggt	atgagttagt	catgaacttt	ggaattcaca	gaccttggtt	tagatttcag	308520
attctccaat	tattagctat	gtcagttagg	gcaaattggc	ttaattcttt	aaactcaatg	308580
tattgatgtg	gaaaaagggg	caataaatcc	tatcattcaa	agttgtaatg	agcagtaaa	308640
tatatgaata	tgtaaaaatg	actaggctag	tccttgacta	tgggacatag	catttatttc	308700
catgtgcact	ttgtgtgagg	gagtgttaata	tccaatttct	gtattatcat	actgcaacta	308760
tgaatatgga	ccaaatata	gatacatctt	taccatttac	agactacttt	tccatgtaca	308820
cacacatagc	cctttgctaa	aatgtgatca	tggaaattagg	atttcataaa	attgaattat	308880
aacattggaa	ggccatgtag	cttaagatta	aaaacataga	atttgagac	agacacaaatg	308940
agctggaatt	ctattctact	tgtttcctgt	cctatcttga	gcaagccact	caacattttc	309000
agtctctttc	tttttcatgt	aaatgtacat	atccatgtca	tagatttgtg	atgtctatta	309060
aggaagatgt	atgaaagcac	tgtcacaaatg	accaataaaa	ggagaggatt	caataaatgg	309120
tcattgtccc	tttaacagatg	tctttaaaaa	cataattaaat	ttgaaacatt	gtttcagagt	309180
gtcttattgg	atgaaatgat	aataactgta	catatgggtg	aagattagt	attggaagtc	309240
aagaattttt	gcttcagctc	cttgaccttg	aggttgaaat	ttaagcctcc	tcattctagt	309300

tttttaaaaa	acgttttctaa	cagaaatata	ctttgtggat	cattatacct	catttctaate	309360
cagctccctc	aagcttaate	tgtatgaact	ttggggaaaa	gataagcact	gtgaataagt	309420
tggcatgctt	gaacattctt	tgaataagta	gaaaatattg	ctctttcatt	ttattgacag	309480
aaagattatc	ttgatagctt	atacttctgt	ttttaatgta	aataataggt	atcaaattct	309540
aaaattgaat	aatggtgacc	aatttaatga	gtgtctgtaa	taaggggaat	ttataaatgt	309600
tttgactaa	ttgcagtttg	taaagttatt	gaacttccct	ttatacttag	atgataattt	309660
attattttatt	tattttattt	tgagatggag	tcccactctg	ctgcccagag	tggagtgcag	309720
tggcccgatc	tcagctcact	gcaacctctg	cctcccaggt	tcaagtgatc	cttctgcttc	309780
agtctcctga	gtatccggga	ttacaggcac	gtgccaaac	gcccagctac	ttttttgtat	309840
ttttagtaga	gatgaggttt	cactatgtta	gccaggatgg	tctccatctc	ttgacctcat	309900
gatctgcccc	cctgggcctc	ccaaagtgtc	gggattacag	gcgtgagtca	ctgcgcccgg	309960
accaattttt	tatttttatt	gacaatgctg	ctttggataa	agaaagttag	tgaacgcctt	310020
ggtcaatgca	actgtggtta	agagaacaga	aatgatcatg	cacaatagca	agatttcagt	310080
taaggttctt	ggttgcaagt	gataaaaaatt	taaagcagga	aaataatcta	ttagaaggct	310140
atggaataga	tgctcagagac	aaatgttaag	gttctgaaat	tgggcaataa	gcaagggaag	310200
aaagacctca	aaggcattaa	cctgcagtgc	ttctgaacgt	gggtctctga	gctgaagttc	310260
caaatctctg	ctccaccact	taatagaact	ttgtgacttc	gggcaagtga	cttaaccttt	310320
gtatgtccca	gtattcagat	cttcaaaatt	agactaccct	tacagctaac	tgagttaaca	310380
cgtgaaagac	acataataga	tcttcaaaaa	tgccagaaaa	aaaataaatg	tggccagaac	310440
aatggccatc	acattatata	ccatggactc	cactggacgt	ggttatggaa	aactaatctc	310500
gctgtccctt	tgaatttgca	tacctccaca	aagttcctga	cctatatattg	attttgcctt	310560
ggttgagagg	agaacagaag	ggatgaaatt	tctttacttt	ttctctaagt	gaaaggtata	310620
attccctgaa	actaagaaga	ggagggattt	ggatgttaaa	cagccacaca	ttttgaaaat	310680
gtccactaca	gttagaataa	ccatatatcc	tagctacott	gagaaagtcc	cagtttacaa	310740
ctgatatccc	acatggttaa	acacaccttc	gtggcagaag	tgccacaatt	tggatgataa	310800
attatgggtca	ctgtgattgt	cattcactcc	cttttgcctg	taaaaatgtgc	atatattctt	310860
cttcttatac	ttactggttc	agtgaaaaaa	gaccttaata	acaaaattaa	attttcattt	310920
ataaaaaaaa	taatgcattt	attcttttca	gcaaaacaaa	caaaaacaca	acaacaacaa	310980
caacaaaaaa	ccagaaagtc	caaaatacta	ttattctgct	atctctaagt	cttggtatct	311040
tgtggactat	tttttgactc	taattttttc	taactctatc	tcagtatttc	atcacctaca	311100
gagttaaat	gcatogtccc	cacaacaata	ccctacatac	ggtatctata	tctattttata	311160
tacatgtata	caaatatata	cataaatata	tatatgcata	tatacatatc	tatatgtata	311220
tagatagatg	tagatatatg	aaaggatat	ctatatacag	taggcctcca	ttacctcaga	311280
gacacatctc	aagacctcca	gtggatgcct	gaaaacgtgg	atagtgccta	accctatata	311340
tactatgtat	aatttttttc	tttgtcacca	tttccacagat	aaaagattca	ttttgctgta	311400
tatcttgaca	acctcagcat	atgatttttt	tccctttott	tatttaagtca	agaacttgca	311460
tcattttctc	taatggaagc	actttacagc	ttctgtttgg	catagccaaa	tggccagtat	311520
cactgctttt	gtgctttgag	gccattatta	agtcacataa	ggatgacttg	gacataagca	311580
ctgaaatagt	caatctgata	gctgggggtg	ctagttagtg	actaatgggt	ggggagcaga	311640
gacagccctg	agacctgaa	caaagtgaat	attctcccg	agtttgatgg	agaaagagca	311700
caaaacttca	tcaccatact	cagaacagca	cacaatttaa	aacttatgaa	ttgtttattc	311760
ctggaatttt	ccacttaata	tttttgagac	atggtttact	gcaggtcatt	gaaacctaaa	311820
aaagcaaaac	tttgaataag	gagaaactac	tctgtgtgtc	tacatgcaca	tgtataagta	311880
gtctctcttt	cacacacata	tacacatata	ggaaaggaag	agaagacatg	tggattctag	311940
tttctcatt	tcataacttg	attatgaggc	caaattgagt	atttatgact	tgctccttc	312000
atgacctcag	acatactccc	tctactttcr	gggaacactt	ctgtaatatg	ctagaatttt	312060
catgcctaaa	gattctaagg	gtttttcttt	ttacatatag	aatgaaaaga	gtctttcatt	312120
aacttccaat	taatgaacat	gctgggtctg	tacaccatct	accacccaga	acgtgatctt	312180
ttcaagatgt	tgtttcccat	ctggcactat	agtgagtttt	ccatgaatca	tccagtattc	312240
cacaagggtc	acacagtaag	aaccttgaat	cctatgggaa	taagctcaat	gctaatactc	312300
gtttgcatta	tattctttgt	ttagtagcaa	tgatatataa	aagttgtggg	tatagagttt	312360
gttgaaaata	tctgaagaag	gtacagacag	caaaggcaga	cacagaagaa	aaatatatca	312420
attttttaac	agcagctggt	ctactgaagg	caaatcactg	cactctccat	gattcaagaa	312480
atcaactacc	acccccccaa	ccccccctc	accacacaca	cacgcagtag	tatttgctag	312540
catcagctct	gctcctctat	cagggcctct	aggttgctgt	cagtatgaat	gaattagaca	312600
ctcaggggtg	ccattagctt	ggcaacctgt	ccattttacc	atgtccctagg	catgtttcta	312660
tcataagac	ctttgtcact	tgacttaata	aggagacaca	cagtggccag	gaaaagagat	312720
atttcatcac	agaacctgct	gccatgtttg	ttagaaggag	aaggagctag	atatgtgact	312780
atgtgtccat	tcttgagccc	tggcacttgg	gttgcttga	tggtcaggaa	cttgggaagg	312840
agacaattga	aatattgatg	acaagaaaat	ttggggaaga	gacatgtggc	tataactctc	312900
agaatgaacc	aaaacttgaa	gatatttttg	gatgttttaa	aaaggatgac	ctcagaatag	312960



gaggattttt	acaattaaaa	gaataggatg	acttgttttg	tgggtgccag	tcagcctctt	313020
tactcagtc	tocctgtcac	ttcccattgg	gctcatgact	aaaatagcca	ttgtgtcagg	313080
gatggagggt	acgcatgggc	tcagctacat	ggacttacac	tcaccaaggc	tgcccttcta	313140
ctatcagtct	accagctgta	gagaccaaca	ctgtgcaccc	aatgtgacac	tattccccag	313200
gggatcaggc	atctacctgg	tggcaggcag	atcacagtgg	acagcctcca	tcattggaaa	313260
ggaagtattt	tgtttctact	ggataagatt	ttggatgtgg	atgtgtctgt	cctgaacaca	313320
atgcttctgc	caaacacgga	cttacagact	gccttatgca	ctgtcgcagt	attcctcacg	313380
gcatggcttc	tgatcaagga	actcacttca	cagaaaaaga	agagtgccaa	tggggccatg	313440
tcattgggatt	caccgggttt	atcatgttcc	ccaccgtcct	gaagcagctg	gcttgataga	313500
atggtaaaat	gaccttttga	agacagctct	agccccagct	aggttaacttc	acagggcggg	313560
ggcgagggtc	tctagatggc	tgtaaatgct	ctaaactaac	tatccgatat	atgggtgtcat	313620
taatttttat	ccaaattcat	cttcttcacc	tgtgtattat	caatagaaaa	tagcttaatt	313680
tgtcaatcta	tcttgggtca	aagtttacca	taaaactttt	catgaatttt	agtttcaaaa	313740
atttacctct	tatgtaatat	tgtgtatata	tttatatagt	aagaaataat	atagtaaaaga	313800
aaatcttaaa	cacatggcta	agtttggcag	aatgtcataa	agatcacatt	gtacaacaaa	313860
tgccagaaat	tatagtgtct	tccataattt	tgtttctttg	agactgattc	cactgggtttt	313920
caaatattgc	tacagttgaa	acaaaattaa	aagaataaaa	actcagaatg	atcacataga	313980
gtggctaaat	taatgatatt	caagttctgt	ttcttcatct	ttataatcac	ggtgttataa	314040
ttgtttattt	caaatcaacc	gctcaaacac	aggaatccat	agcaccaaaa	atttttgata	314100
cagaaacttc	actgagggca	agcctaaatg	atcccaaatc	tctatgaacc	ttttcttgaa	314160
atgaatgcat	gtaactgagt	ccgtgtttgt	caaggccatt	tgtaaaactg	gaagttctct	314220
aatttaacat	ccatgtaaaa	tacaatat	ataaaacaac	ttgtaataaa	tgtatgttaa	314280
tttgatgtta	catatatatg	tatatatnca	attacatata	tatatatata	attaaattca	314340
agagaaacca	agtgtattgt	taggactgag	accaataaaa	tatttaataa	aataattttg	314400
gagggaggac	ttgtaatcac	gaactggggg	gaatatccac	gggcaaacca	tgaaccagat	314460
tcaggacaat	gtcgttgtcc	tctgaatccc	aggtgaagtc	actttgactc	ttaaatggca	314520
gaaggactct	gggttttcaa	aaccataaag	cctgttatct	aacgtatttt	cgtttccacca	314580
gtgcacagct	cctcttgttt	atggatgcag	atgcctttcc	agaaggctag	aaattataga	314640
aataacattt	ggtattttca	gagtctgtat	atggaagtac	ctacctacca	tgtttttatt	314700
taaagaattt	tgaatatgtg	atttgactca	aatttgggaa	ttgtgtgagg	aacaaaacta	314760
ttttggtggc	ttcaattagg	tctatatgca	atagaacgtg	aatacttttc	aaaaacagaa	314820
atagaacagt	gcctacgtga	gctagcacca	attatgggtcc	ttggtactcc	atggcctatt	314880
tacagctaca	ttaatccatg	agtcttataa	ttcaaattgg	tccagtggct	ttgatgttca	314940
ttagggacac	aacagccctt	tgccctccct	ggagaagtgc	tgtgttttgg	gttctaacaa	315000
cataaatact	ccccagtccc	agtgttttta	tggagcccaa	ttcaatgtac	gatctccccg	315060
tggcatcacc	tacctataat	gattaaggat	ttgggagatt	catttagtga	attttatcaa	315120
gataacaatc	ttaacaaaca	ctgtagcata	tgcacactcc	aaaggctttc	cctctttcta	315180
cgttaaatcc	aatggattcc	tgggactgtg	tctcttatag	gcttttactg	ggcctaaatt	315240
ttagccaata	agcttgtcaa	acagctagaa	acacatttaa	ctcttcaagc	ttacctactc	315300
tatccatgct	cttcgggtaa	gtgtccactt	tgaatttcac	cccaatgcaa	tcttttatct	315360
tgggttgatg	cacaccttca	gaacctattc	ctaagaaatt	cttgcatata	ttgttggcat	315420
aaatgtcttg	caattctgtt	gttatatcat	ttttttctcg	agtatatttg	acccctattt	315480
aatgcttagc	tottgtctca	gttctaggac	tgcaggcaat	gaagactggg	taagtagggg	315540
ttaggatgag	aattatcttc	tcagtctgag	gtaacttcc	catatgtaaa	ggctactaag	315600
gcagggttgg	aaatgggttt	atatatgcat	ttccgaatat	cctaataata	tttggctgct	315660
tacattagta	aaaatcatta	aagttgcttt	cactcaaaga	ctgtaagata	taaagaaaag	315720
ttatttataa	tgtaaagcat	tgaacagaaa	acaatcgga	actaatcaat	gaaatgggta	315780
tatataggag	gattccaaaa	attcggagac	tctaaccagt	gaatccagcc	ttgctagtac	315840
atccctgcag	ccagatacag	ttgcacttgc	tggaccacta	caacatcggg	tgccaatcta	315900
accactgaat	ccaacctcct	tgtacatcac	tgcaccagga	tgcagttgca	cttgctggac	315960
cactacaaca	tgggtgcca	atttaatact	ttcttctttt	cactctctac	cccrgtttca	316020
gtatccttgg	cagtatgcta	tgttcaactg	actttgagat	gtatgattct	attttgacac	316080
caacaggctt	agaaaagagt	aaactgaaca	ctctggcttt	tgtagggaga	agtagaactc	316140
cgtgtagcac	gaagtcttat	gcaaaagtag	gaagcagtg	aggaggattg	gcagacccaa	316200
aaatgaaaag	tgaacaaat	ttatatacag	ccagataa	ttgttataag	tcaataactc	316260
agatttttta	aatctcccat	ctccaaggca	atttaggcaa	cagcaagtat	cttataaaat	316320
tctctttttt	tagaaaatat	agatgaatta	tttaagatgg	attatcatta	tttattgtgt	316380
atttagtata	ggtttattta	tctcacactt	ctttgtcttt	ttccattatt	ttcagctgtt	316440
caaaaattct	gcatttgtaa	caacttcttc	aatatgtaaa	tttaaaaatc	cctcaatatt	316500
tgcactctta	tcttgaagt	tttaccatct	tactctaaag	taacttgaac	accacatgca	316560
ttattgttgg	agttttaaaa	cattagtatt	tatctgatcc	taattgttta	ttgtaattgt	316620



atagattttt tcactaattt tgggaagggt aactattggt tcatagatgc aataaatcct 316680  
ggaaacttgt ttgcaatact tttaaaagca gttttcaaaa tttctcctct atcacacagg 316740  
agacacactg taatgactac tatttcctct gatttccttt aggtgcgttt cttttgactg 316800  
tattttcatc agtctcatca tttttctgtg tcattccattt gogctgggac ttgtgcctac 316860  
tgatggtttt caagtgtggc tgcaatttgg ttcctctgaa ggtattttta aatatgctga 316920  
tgcactagta ccattcccaa accaactgaa acacattctt caatggcaag gtcattggtg 316980  
tggattttta aacaaaaatc ttcctaagtt attttaatat acaactacaa catattcact 317040  
ggtatagata atattgaata caaaaaaac ctatatgatg actcgttcct tctgtgagag 317100  
agcaaagtcc acaattttaa aaaatttata atactgttaa ctacagaggga tcagaccagg 317160  
tggaaagcac gaatatgcac tctgcataat gaagcttatt tccagtgtac aaatctacag 317220  
caagcagaaa aggatacttg gtgcagtttg ctctagatac acagctaaca cagcgagcca 317280  
tcagtgggg cctccaccta gaggaactga ttacaggaag aggaagtgtt gatttcaacc 317340  
ttgttaggct ttccttccag ttagtaattg tttcaagctg aatggatgac acaacatgaa 317400  
aaattcaact acttaagaat agcatgcgcc aagtatatct tataactttc tgttaagatt 317460  
catttcaact aataattttc tatataaatc tctattaatt taagtttctg aataactgta 317520  
gggtcaacaa catatttggg actctcctac tttgagcttg gttgaaccac ataatttgaa 317580  
agagatatt ttataaagtt aaaaactga taactttaaa gactgataaa tttctgtttg 317640  
cattcagaat atagtggcag aaaaaaatat tcttttctaa attataatga ttaaattgtt 317700  
ttgatggtat tatgttgcta taaataattt cctagagata ttaaggataa cttacataag 317760  
atttattcac atttatgata ttttaggctg accttaaaaa ttatttatga atttaattgc 317820  
ttatgcata gtgataaag ctaattataa tatttacttg gotttaataa atatttccag 317880  
ttgcaatttg ataggctaac tgtatgtaaa aatttcaataa tattgtctag aatagttttt 317940  
ctttattctg atccttttta gttatagtct tatagatatg gccttagttc aaaagcacta 318000  
tacactcgtg aaacattttc aagtcaatgt aaaatattgt aatgtgatgt cagaataaaa 318060  
tctttttgaa gatagagatt ttcattggtt ttttccaaat cagagttaa agtagtcaat 318120  
ctcatgattt gtacagactg aaaagtata gaatgaatat ttaacataa aaataactta 318180  
aatgaatato cttagaatac aaatcaactg tatattcact cgaggcactt aaaatattct 318240  
atatgccaaa aactgcatac ctctttccaa ttttaatac tgccttctct ctctctctct 318300  
tagtttacaa atatttctca atattttaa ttttagaact gaatcatggt ggggaaatga 318360  
agtctttggg gaaataattt attatgaaaa atttttatgt agttatttaa aaaattctaa 318420  
ctcagtttaa caatattata tggaaatgcc atagtttgaa aaattgtcaa ggttaagata 318480  
aaacaaaaaa atatttttga cttttttttt cttttctttt attattatac ttttaagttt 318540  
agggtacatg tgcacattgt gcaggttatt tacatatgta tacatgggcc atgctggtgt 318600  
cctgcacca ttaacttgct atttagcatt aggtatatct cccaatgcta tccctcccc 318660  
ctccccccac ccacaacan gtccccagag tgtgatgttc cccttctgt gtccacgtga 318720  
tctcattggt ncaattccca cctatgagtg agaatatgag gtgtttggtt tttgttctt 318780  
gtgatagttt actgagaatg atgatttcca atttcatcca tgtccctaca aaggacatga 318840  
actcatcatt ttttatggct gcatagtatt ccatagtgta tatgtgccac attttcttaa 318900  
tccagtctat cattgttgga catttggtgt ggttccaagt ctttgcatt gtgaataatg 318960  
ccacaataaa catacgtgtg catgtgtctt tacaacaaca tgatttatag tctttgggt 319020  
atataccag taatgggatg gctgggtcaa atggtatttc tagttctaga tccctgagga 319080  
atcgccacac tgacttccac aatggttgaa ctagtttaca gtcccaccaa cagtgtaaaa 319140  
gtgttccat tctccacat cctctccagc acctgttggt tcttgacttt ttaatgattg 319200  
ccattctaac tgggtgtgaga tggatatctca ttgtggtttt gatttgcatt tctctgatgg 319260  
ccagtgtgta tgagcatttt ttcattgtgc ttttggctgc ataaatgtct tcttttgaga 319320  
agtgtctgtt catatccttt gccactttt tcatggggtt gtctgttttt tctttgtaa 319380  
tttgtttgag ttcattgtag attctggata ttagcccttt gtcagatgag taggttgaga 319440  
aaattttctc ccattttgta ggttccatgt tcatctgat ggtagtctt tttgctgtgc 319500  
agaagctctt taatttagtt agatcccat tgtcaatttt gtcttttgtt gccattgctt 319560  
ttggtattta gacgtgaagt ccttgcccat gcctatgtcc tgaatggt 319608

&lt;210&gt; 2

&lt;211&gt; 2226

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; 5' UTR

&lt;222&gt; 1..95

&lt;220&gt;

<221> CDS  
<222> 96..167

<220>  
<221> 3'UTR  
<222> 168..2226

<220>  
<221> polyA\_signal  
<222> 2200..2205

<220>  
<221> allele  
<222> 132  
<223> 8-130-143 : polymorphic base A or G

<220>  
<221> allele  
<222> 132  
<223> 8-126-286 : polymorphic base A or G

<400> 2  
ttgcaggctcc acattcggtc tagaggctgc ctgcggctcct tagaggcctc ataggcaagg 60  
aactcatagt ttcttgccat gaagactttt ccaac atg gct gct tac ttc ttc 113  
Met Ala Ala Tyr Phe Phe  
1 5  
aag cct gta agg aga tcc rgt ctc agt tgg cca ggt ctt gtg aaa gga 161  
Lys Pro Val Arg Ser Xaa Leu Ser Trp Pro Gly Leu Val Lys Gly  
10 15 20  
aaa taa cgcatttata aaagccagct cagatgtttac ttccaaccaa caaatttctc 217  
Lys \*  
ggacacacta agtaatagag ttaatcacac cgtcctgaac attcaccctg ggagctctgt 277  
gctgccatgc aaatgtcaca ttgcactgag gccgccatgc tgtaaaaaac cccaagctac 337  
ctacctgaaa agctggagag aatcagcacc catcagcttt tccagcattt tgggtcctcc 397  
cagctaaatc atcggcattg tgaagcagag atgagacctc ctgccatgcc ctttccaaat 457  
cctggaccaa cagaatcatc atatacacta ggttttgtgc atctctaaga tgatcatctt 517  
ggtgtggggt cctgatcttc ccagtgactt agcagcacag taatttggct gaaatgttga 577  
tgagttcaga caagaagaag gtgatgatga caagagagcg tcacggggaa attctgcctt 637  
gaggaaggaa tagtttgagg ctggagagat gtgctacccc cgccactctt tctcaaatgg 697  
aaagacaacc tgtaggcaca gaaatcattg ctctgcttac actgaaatct cctgctgact 757  
ttgagagctg agcattggat tttgatggga gagtgaact tagtgaagaa atgatttgaa 817  
cctttgttga tctaatactt ctgtgataca gttatctaaa cagaatgagt cttaaagtcag 877  
tataaagacc tgtgtctctc catgttttgc tgcttataat tgcagaggaa cattttgaga 937  
tagaaaattt tatccctctt tcaagtgact ttttataaaa cataacctt taagagaggt 997  
cattcttttg aaaaattgtc tgaatacaaa aaagtaactg aacagcaagg gattgtagcc 1057  
ataatgttct tgaccacttt agattgacaa atagtatctc aaaattgac aggttctgga 1117  
gaatctattg aggttgatg gagtaaccaa gtgatggaag aggttggaag tgaataattt 1177  
tatgtggtat tatgtgattt tttttggtgg gggtagtgct tggagctgag catgatggag 1237  
cctagatggg gcagccttct aatgctccca gtatttcgca aatcatatgt cagtcacaaa 1297  
ttgtaatgac aaagctcaca gcagctgcct acgccaggtt tggtgccagc tctaattaga 1357  
gcattgatca tttctctccc ctcatgaat gcacatgagt gcacattttc atataaagg 1417  
gcaaaggaca tttcaaagtt tggatggaat ctttcctact tagaagtaat aaaaaacaat 1477  
actttgaaat cttaccaaat gttaaacaaa tagcaacaat ttttccattt gaagatgtgt 1537  
gtcttaagggt ctggttgatc tttcaaattt ctattttgga catggaaact acagatgaaa 1597  
tttaacttac tcaaagttac acagaactcc tggctgatct cttttactta attcttctaa 1657  
aatctaaagt aagaataatt ttgaaatttt atttgtaatg cttaaataac aaaagttact 1717  
ttaatatgtt acatgtttga agtttcattt atacttaata tctaatttaa cgcttgatga 1777  
aatattattc agtatttttt tttcttttct aaaatgactt gttcagggtt ccttttgagg 1837  
ttttaagtga gaaaaactct atttcataat attgttacca tgcctcatcca ataacacttc 1897  
atagatgggt agccatctct aaaatttaaat tttgtagttg aggagcattg aaaatatcac 1957  
ctatatattg atttctgttt aacacttcaa aatgcatttt atcataattt aaaatttgat 2017

ttaaatacaa acctctttat agaaaacctg ataaatgtag ttttatttgt ttgcttgctt 2077  
 ttttttgcag ggaggatat tcccttgata ccactctgtc tttatgactc atcggttatt 2137  
 tataaatggt atcatcggtt ttctgatatt taaataagtt tcagaattaa tacatttaac 2197  
 tgaataaatc aaccaataat gtcaaacctt 2226

<210> 3

<211> 2431

<212> DNA

<213> Homo sapiens

<220>

<221> 5'UTR

<222> 1..95

<220>

<221> CDS

<222> 96..167

<220>

<221> 3'UTR

<222> 168..2431

<220>

<221> polyA\_signal

<222> 2405..2410

<220>

<221> allele

<222> 132

<223> 8-130-143 : polymorphic base A or G

<220>

<221> allele

<222> 132

<223> 8-126-286 : polymorphic base A or G

<220>

<221> allele

<222> 328

<223> 8-155-258 : polymorphic base C or T

<400> 3

ttgcaggctc acattcggtc tagaggctgc ctgcggctct tagaggcctc ataggcaagg 60  
 aactcatagt ttcttgccat gaagactttt ccaac atg gct gct tac ttc ttc 113

Met Ala Ala Tyr Phe Phe  
 1 5

aag cct gta agg aga tcc rgt ctc agt tgg cca ggt ctt gtg aaa gga 161  
 Lys Pro Val Arg Arg Ser Xaa Leu Ser Trp Pro Gly Leu Val Lys Gly

10 15 20

aaa taa cgcatttata aaagccagct cagatgttac ttccaaccaa caaatttcct 217  
 Lys \*

ggacacacta agtaatagag ttaatcacac cgtccttgaa ggtctccctg ctgtgcttag 277

agagggggat gctggaagaa gagggtgacg agaagcaaca ttgctgactc yaaagacaga 337

gggcaggcca cgagccaagg catgcaggag gcctccagaa gctggggcag gcaagagaat 397

ggactcagcc ctggagcctc cagaaggcat gcagccctgc ccattccttg cttttagcca 457

ggaacattca ccctgggagc tctgtgctgc catgcaaatg tcacattgca ctgaggccgc 517

catgctgtaa aaaaccccaa gotacctacc tgaaaagctg gagagaatca gcacccatca 577

gcttttccag cattttgggt cctcccagct aaatcatcgg cattgtgaag cagagatgag 637

acctcctgcc atgccccttc caaatcctgg accaacagaa tcatcatata cactagggtt 697

tgtgcatctc taagatgac atcttggtgt ggggtcctga tcttcccagt gacttagcag 757

cacagtaatt tggctgaaat gttgatgagt tcagacaaga agaaggtgat gatgacaaga 817

```

gagcgtcacg gggaaattct gccctgagga aggaatagtt tgaggctgga gagatgtgct 877
acccccgcca ctctttctca aatggaaaga caacctgtag gcacagaaat cattgctctg 937
cttacctga aatctctgct tgactttgag agctgagcat tggattttga tgggagagt 997
aaacttagtg aagaaatgat ttgaaccttt gttgatctaa tacttctgtg atacagttat 1057
ctaaacagaa tgagtctaaa gtcagtataa agacctgtgt ctctccatgt tttgctgctt 1117
ataattgcag aggaacattt tgagatagaa aattttatcc ctctttcaag tgacttttta 1177
taaaacataa ccttttaaga gaggtcattc ttttgaaaaa ttgtctgaat acaaaaaagt 1237
aactgaacag caagggattg tagccataat gttcttgacc acttttagatt gacaaatagt 1297
atctcaaaat tgatcaggtt ctggagaatc tattgaggtt gtatggagta accaagtgat 1357
ggaagaggtt ggaagtgaat aattttatgt ggtattatgt gatttttttt ggtgggggta 1417
gtgtctggag ctgagcatga tggagcctag atggggcagc cttctaatagc tcccagttat 1477
tcgcaaatca tatgtcacgt caaaattgta atgacaaagc tcacagcagc tgcctacgcc 1537
aggtttggtg ccagctctaa tttagagcatt gatcatttct ctccccttca tgaatgcaca 1597
tgagtgcaca ttttcatata taagggcaaa ggacatttca aagtttggtt ggaatctttc 1657
ctacttagaa gtaataaaaa acaatacttt gaaatcttac caaatgttaa acaaatagca 1717
acaatttttc catttgaaaga tgtgtgtctt aaggtctgtt ggatctttca aatttctatt 1777
ttggacatgg aaactacaga tgaaatttaa cttactcaaa gttacacaga actcctggct 1837
gatctctttt acttaattct tctaaaatct aaagtaagaa taattttgaa attttattgg 1897
taatgcttaa ataacaaaag ttactttaat atgttacatg tttgaagttt catttatact 1957
taatattctaa tttaacgctt gtgaaaatat tattcagtat ttttttttct tttctaaaat 2017
gacttgttca gggctccttt tgggatttta agtgagaaaa actctatttc ataatttgt 2077
taccatgctc atccaataac acttcataga tggttagcca tctctaaaaat taaattttgt 2137
agttgaggag cattgaaaat atcacctata tttggatttc tgtttaacac ttcaaaatgc 2197
attttatcat aattttaaatt ttgatttaaa tacaaacctc tttatagaaa acctgataaa 2257
tgtagtttta tttgtttgct tgcttttttt tgcagggagg gatattccct tgataccact 2317
ctgtctttat gactcatcgg ttatttataa atgttatcat cgtttttctg atttttaaat 2377
aagtttcaga attaatacat ttaactgaat aaatcaacca ataatgtcaa actt 2431

```

&lt;210&gt; 4

&lt;211&gt; 2215

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; 5'UTR

&lt;222&gt; 1..95

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 96..311

&lt;220&gt;

&lt;221&gt; 3'UTR

&lt;222&gt; 312..2215

&lt;220&gt;

&lt;221&gt; polyA signal

&lt;222&gt; 2189..2194

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 132

&lt;223&gt; 8-130-143 : polymorphic base A or G

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 132

&lt;223&gt; 8-126-286 : polymorphic base A or G

&lt;400&gt; 4

ttgcaggtcc acattcggtc tagaggctgc ctgcggctct tagaggcctc ataggcaagg 60

aactcatagt ttcttgccat gaagactttt ccaac atg gct gct tac ttc ttc 113  
Met Ala Ala Tyr Phe Phe  
1 5

aag cct gta agg aga tcc rgt ctc agt tgg cca ggt ctt gtg aaa ggt 161  
Lys Pro Val Arg Arg Ser Xaa Leu Ser Trp Pro Gly Leu Val Lys Gly  
10 15 20

aac ctc atc acc gga gtc aag tgt cat cac ctt tcc cat gtt cca ttg 209  
Asn Leu Ile Thr Gly Val Lys Cys His His Leu Ser His Val Pro Leu  
25 30 35

gtt aga agt ccc agc tcc tgc cca cat tca agg gaa cat tca ccc tgg 257  
Val Arg Ser Pro Ser Ser Cys Pro His Ser Arg Glu His Ser Pro Trp  
40 45 50

gag ctc tgt gct gcc atg caa atg tca cat tgc act gag gcc gcc atg 305  
Glu Leu Cys Ala Ala Met Gln Met Ser His Cys Thr Glu Ala Ala Met  
55 60 65 70

ctg taa aaaaccccaa gctacctacc tgaaaagctg gagagaatca gcacccatca 361  
Leu \*

gcttttccag ctttttgggt cctcccagct aaatcatogg cattgtgaag cagagatgag 421  
acctcctgcc atgccctttc caaatcctgg accaacagaa tcatcatata cactagggtt 481  
tgtgcatctc taagatgac atcttggtgt ggggtcctga tcttcccagt gacttagcag 541  
cacagtaatt tggtgaaat gttgatgagt tcagacaaga agaaggatgat gatgacaaga 601  
gagcgtcacg gggaaattct gccctgagga aggaatagtt tgaggctgga gagatgtgct 661  
acccccgcc ctctttctca aatggaaaga caacctgtag gcacagaaat cattgctctg 721  
cttacctga aatctcctgc tgactttgag agctgagcat tggattttga tgggagagt 781  
aaacttagtg aagaaatgat ttgaaccttt gttgatctaa tacttctgtg atacagttat 841  
ctaaacagaa tgagtctaaa gtcagtataa agacctgtgt ctctccatgt tttgctgctt 901  
ataattgcag aggaacattt tgagatagaa aattttatcc ctctttcaag tgacttttta 961  
taaaacataa ccttttaaga gaggtcattc ttttgaaaaa ttgtctgaat acaaaaaagt 1021  
aactgaacag caagggattg tagccataat gttcttgacc acttttagatt gacaaatagt 1081  
atctcaaaa tgatcagggt ctggagaatc tattgaggt gtatggagta accaagtgat 1141  
ggaagagggt ggaagtgaat aattttatgt ggtattatgt gattttttt ggtgggggta 1201  
gtgtctggag ctgagcatga tggagcctag atggggcagc ctctaatgc tcccagtatt 1261  
tcgcaaatca tatgtcacgt caaaattgta atgacaaagc tcacagcagc tgccctacgcc 1321  
aggtttgggt ccagctctaa tttagagcatt gatcatttct ctccccttca tgaatgcaca 1381  
tgagtgcaca ttttcatata taagggcaaa ggacatttca aagtttggat ggaatctttc 1441  
ctacttagaa gtaataaaaa acaatacttt gaaatcttac caaatgttaa acaaatagca 1501  
acaatttttc catttgaaga tgtgtgtctt aaggctgtgt ggatctttca aatttctatt 1561  
ttggacatgg aaactacaga tgaaatttaa cttactcaa gttacacaga actcctggct 1621  
gatctctttt acttaattct tctaaaatct aaagtaagaa taattttgaa atttttattg 1681  
taatgcttaa ataacaaaag ttactttaat atgttacatg tttgaagttt catttatact 1741  
taatatctaa ttttaacgctt gtgaaaatat tattcagtat ttttttttct tttctaaaat 1801  
gacttgttca gggctccttt tgggatttta agtgagaaaa actctatttc ataattattg 1861  
taccatgctc atccaataac acttcataga tggtagcca tctctaaaat taaattttgt 1921  
agttgaggag cattgaaaat atcacctata tttggatttc tgtttaacac ttcaaaatgc 1981  
attttatcat aatttaaaat ttgatttaaa tacaacctc tttatagaaa acctgataaa 2041  
tgtagtttta tttgtttgct tgcttttttt tgcaggagg gatattccct tgataccact 2101  
ctgtctttat gactcatcgg ttatttataa atgttatcat cgtttttctg attttttaaa 2161  
aagtttcaga attaatacat ttaactgaat aaatcaacca ataatgtcaa actt 2215

<210> 5  
<211> 2132  
<212> DNA  
<213> Homo sapiens

<220>  
<221> 5'UTR  
<222> 1..95

<220>  
<221> CDS  
<222> 96..212

&lt;220&gt;

&lt;221&gt; 3'UTR

&lt;222&gt; 213..2132

&lt;220&gt;

&lt;221&gt; polyA\_signal

&lt;222&gt; 2106..2111

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 132

&lt;223&gt; 8-130-143 : polymorphic base A or G

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 132

&lt;223&gt; 8-126-286 : polymorphic base A or G

&lt;400&gt; 5

ttgcaggtcc acattcggtc tagaggctgc ctgcggtcct tagaggcctc ataggcaagg 60

aactcatagt ttcttgccat gaagactttt ccaac atg gct gct tac ttc ttc 113

Met Ala Ala Tyr Phe Phe

1

5

aag cct gta agg aga tcc rgt ctc agt tgg cca ggt ctt gtg aaa gga 161

Lys Pro Val Arg Arg Ser Xaa Leu Ser Trp Pro Gly Leu Val Lys Gly

10

15

20

aca ttc acc ctg gga gct ctg tgc tgc cat gca aat gtc aca ttg cac 209

Thr Phe Thr Leu Gly Ala Leu Cys Cys His Ala Asn Val Thr Leu His

25

30

35

tga ggccgccatg ctgtataaaaa ccccaagcta cctacctgaa aagctggaga 262

\*

gaatcagcac ccatcagctt ttccagcatt ttgggtcctc ccagctaaat catcggcatt 322

gtgaagcaga gatgagacct cctgccatgc cctttccaaa tcctggacca acagaatcat 382

catatacact aggttttgtg catctctaag atgatcatct tgggtggtggg toctgatctt 442

cccagtgaact tagcagcaca gtaatttggc tgaaatgttg atgagttcag acaagaagaa 502

ggtgatgatg acaagagagc gtcacgggga aattctgccc tgaggaagga atagtttgag 562

gctggagaga tgtgctaccc ccgccactct ttctcaaatg gaaagacaac ctgtaggcac 622

agaaatcatt gctctgctta cactgaaatc tcctgctgac tttgagagct gagcattgga 682

ttttgatggg agagtgaaac ttagtgaaga aatgatttga acotttgttg atctaatact 742

tctgtgatac agttatctaa acagaatgag tctaaagtca gtataaagac ctgtgtctct 802

ccatgttttg ctgcttataa ttgcagagga acattttgag atagaaaatt ttatccctct 862

ttcaagtgaac tttttataaa acataacctt ttaagagagg tcattctttt gaaaaattgt 922

ctgaatacaa aaaagtaact gaacagcaag ggattgtagc cataatgttc ttgaccactt 982

tagattgaca aatagtatct caaaattgat caggttctgg agaattctatt gaggttgat 1042

ggagtaacca agtgatggaa gaggttgga gtgaataatt ttatgtggta ttatgtgatt 1102

tttttgggtg ggggtagtgt ctggagctga gcatgatgga gcctagatgg ggcagccttc 1162

taatgctccc agtatttcgc aaatcatatg tcacgtcaaa attgtaatga caaagctcac 1222

agcagctgcc tacgccaggt ttggtgccag ctctaattag agcattgatc atttctctcc 1282

ccttcagtaa tgcacatgag tgcacatttt catatataag ggcaaaggac atttcaaagt 1342

ttggatggaa tctttcctac ttagaagtaa taaaaaacia tactttgaaa tcttaccaaa 1402

tgtaaacia atagcaacia tttttccatt tgaagatgtg tgtcttaagg tctgttgat 1462

ctttcaaatt tctatttttg acatggaaac tacagatgaa atttaactta ctcaaagtta 1522

cacagaactc ctggctgatc tcttttactt aattcttcta aaatctaaag taagaataat 1582

tttgaattt tatttgtaat gcttaaataa caaaagttac ttaatatgt tacatgtttg 1642

aagtttctatt tatacttaatt atctaattta acgcttgatg aaataattatt cagtattttt 1702

ttttcttttc taaaatgact tgttcagggc tccttttggg attttaagt agaaaaactc 1762

tatttcataa tattgttacc atgctcatcc aataacactt catagatggg tagccatctc 1822

taaaattaaa tttgtagtt gaggagcatt gaaaatatca cctatatattg gatttctgtt 1882

taacacttca aaatgcattt tatcataatt taaaatttga tttaaataca aacctcttta 1942

tagaaaacct gataaatgta gttttatttg ttgtcttgct tttttttgca gggagggata 2002

ttcccttgat accactctgt ctttatgact catcggttat ttataaatgt tatcatcggt 2062  
 ttcttgattt ttaaataagt ttcagaatta atacatttaa ctgaataaat caaccaataa 2122  
 tgtcaaactt 2132

<210> 6  
 <211> 1981  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> 5'UTR  
 <222> 1..95

<220>  
 <221> CDS  
 <222> 96..260

<220>  
 <221> 3'UTR  
 <222> 261..1981

<220>  
 <221> polyA\_signal  
 <222> 1955..1960

<220>  
 <221> allele  
 <222> 132  
 <223> 8-130-143 : polymorphic base A or G

<220>  
 <221> allele  
 <222> 132  
 <223> 8-126-286 : polymorphic base A or G

<400> 6  
 ttgcaggctc acattcgctc tagaggctgc ctgcggctct tagaggcctc ataggcaagg 60  
 aactcatagt ttcttgccat gaagactttt ccaac atg gct gct tac ttc ttc 113  
 Met Ala Ala Tyr Phe Phe  
 1 5  
 aag cct gta agg aga tcc rgt ctc agt tgg cca ggt ctt gtg aaa ggt 161  
 Lys Pro Val Arg Arg Ser Xaa Leu Ser Trp Pro Gly Leu Val Lys Gly  
 10 15 20  
 aac ctc atc acc gga gtc aag tgt cat cac ctt tcc cat gtt cca ttg 209  
 Asn Leu Ile Thr Gly Val Lys Cys His His Leu Ser His Val Pro Leu  
 25 30 35  
 gtt aga agt ccc agc tcc tgc cca cat tca agg ggt ttt gtg cat ctc 257  
 Val Arg Ser Pro Ser Ser Cys Pro His Ser Arg Gly Phe Val His Leu  
 40 45 50  
 taa gatgatcctc ttggtgtggg gtctctgatct tccagtgac ttagcagcac 310  
 \*

55  
 agtaatttgg ctgaaatggt gatgagttca gacaagaaga aggtgatgat gacaagagag 370  
 cgtcacgggg aaattctgcc ctgaggaagg aatagtttga ggctggagag atgtgctacc 430  
 cccgccactc ttctcaaat ggaaagacaa cctgtaggca cagaaatcat tgctctgctt 490  
 aactgaaat ctcctgctga ctttgagagc tgagcattgg attttgatgg gagagtga 550  
 cttagtgaag aaatgatttg aacctttgtt gatctaatac ttctgtgata cagttatcta 610  
 aacagaatga gtctaaagtc agtataaaga cctgtgtctc tccatgtttt gctgcttata 670  
 attgcagagg aacattttga gatagaaaat tttatccctc tttcaagtga ctttttataa 730  
 aacataacct tttaagagag gtcattcttt tgaaaaattg tctgaataca aaaaagtaac 790  
 tgaacagcaa gggattgtag ccataatggt cttgaccact ttagattgac aaatagtatc 850

```

tcaaaattga tcaggttctg gagaatctat tgagggttgta tggagtaacc aagtgatgga 910
agagggttga agtgaataat tttatgtggt attatgtgat tttttttggt gggggtagt 970
tctggagctg agcatgatgg agcctagatg gggcagcctt ctaatgctcc cagtatttcg 1030
caaatcatat gtcacgtcaa aattgtaatg acaaagctca cagcagctgc ctacgccagg 1090
tttggtgcca gctctaatta gagcattgat catttctctc ccttctcatga atgcacatga 1150
gtgcacattt tcatatataa gggcaaagga catttcaaag tttggatgga atctttccta 1210
cttagaagta ataaaaaaca atactttgaa atcttaccaa atgttaaaca aatagcaaca 1270
atttttccat ttgaagatgt gtgtcttaag gtctgttgga tctttcaaatt ttctattttg 1330
gacatggaaa ctacagatga aatttaactt actcaaagtt acacagaact cctggctgat 1390
ctcttttact taattcttct aaaatctaaa gtaagaataa ttttgaaatt ttattggtaa 1450
tgcttaataa acaaaagtta ctttaatatg ttacatgttt gaagtttcat ttatacttaa 1510
tatctaattt aacgcttggtg aaaatattat tcagtatttt tttttctttt ctaaaatgac 1570
ttgttcaggg ctctttttgg gattttaagt gagaaaaact ctatttcata atattgttac 1630
catgctcaco caataacact tcatagatgg tttagccatct ctaaaattaa attttgtagt 1690
tgaggagcat tgaaaatata acctatattt ggattttctgt ttaacacttc aaaatgcatt 1750
ttatcataat ttaaaatttg atttaaatac aaacctcttt atagaaaacc tgataaatgt 1810
agttttattt gtttgcttgc ttttttttgc agggagggat attcccttga taccactctg 1870
tctttatgac tcatcggtta tttataaatg ttatcatcgt ttttctgatt tttaaataag 1930
tttcagaatt aatacattta actgaataaa tcaaccaata atgtcaaact t 1981

```

&lt;210&gt; 7

&lt;211&gt; 1992

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; 5'UTR

&lt;222&gt; 1..95

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 96..167

&lt;220&gt;

&lt;221&gt; 3'UTR

&lt;222&gt; 168..1992

&lt;220&gt;

&lt;221&gt; polyA\_signal

&lt;222&gt; 1966..1971

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 132

&lt;223&gt; 8-130-143 : polymorphic base A or G

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 132

&lt;223&gt; 8-126-286 : polymorphic base A or G

&lt;400&gt; 7

```

ttgcaggctc acattcgttc tagaggctgc ctgcggtcct tagaggcctc ataggcaagg 60
aactcatagt ttcttgccat gaagactttt ccaac atg gct gct tac ttc ttc 113
                               Met Ala Ala Tyr Phe Phe
                               1           5
aag cct gta agg aga tcc rgt ctc agt tgg cca ggt ctt gtg aaa gga 161
Lys Pro Val Arg Arg Ser Xaa Leu Ser Trp Pro Gly Leu Val Lys Gly
                10                15                20
aaa taa cgcatttata aaagccagct cagatgtttac ttccaaccaa caaatttcct 217
Lys *

```



ggacacacta	agtaatagag	ttaatcacac	cgctctggtt	ttgtgcatct	ctaagatgat	277
catcttggtg	tggggctcctg	atcttccag	tgacttagca	gcacagtaat	ttggctgaaa	337
tggtgatgag	ttcagacaag	aagaagggtga	tgatgacaag	agagcgtcac	ggggaaattc	397
tgcctgagg	aaggaatagt	ttgaggctgg	agagatgtgc	tacccccgcc	actctttctc	457
aaatggaaag	acaacctgta	ggcacagaaa	tcatgtctct	gcttacactg	aaatctcctg	517
ctgactttga	gagctgagca	ttggattttg	atgggagagt	gaaacttagt	gaagaaatga	577
tttgaacctt	tggtgatcta	atacttctgt	gatacagtta	tctaaacaga	atgagtctaa	637
agtcagtata	aagacctgtg	tctctccatg	ttttgctgct	tataattgca	gaggaacatt	697
ttgagataga	aaattttatc	cctctttcaa	gtgacttttt	ataaaacata	accttttaag	757
agaggtcatt	cttttgaaaa	attgtctgaa	tacaaaaaag	taactgaaca	gcaagggtatt	817
gtagccataa	tgttcttgac	cactttagat	tgacaaatag	tatctcaaaa	ttgatcagggt	877
tctgggagaat	ctattgaggt	tgtatggagt	aaccaagtga	tggaagaggt	tggaagtga	937
taattttatg	tggtattatg	tgattttttt	tggtgggggt	agtgtctgga	gctgagcatg	997
atggagccta	gatggggcag	ccttctaattg	ctcccagtat	ttcgcaaatc	atatgtcacg	1057
tcaaaattgt	aatgacaaa	ctcacagcag	ctgcctacgc	caggtttggt	gccagctcta	1117
attagagcat	tgatcatttc	tctccccttc	atgaatgcac	atgagtgcac	attttcatat	1177
ataagggcaa	aggacatttc	aaagtgtgga	tggaatcttt	cctacttaga	agtaataaaa	1237
aacaatactt	tgaaatctta	ccaaatgtta	aacaaatagc	aacaattttt	ccatttgaag	1297
atgtgtgtct	taagggtctgt	tggtatctttc	aaatttctat	tttgacatg	gaaactacag	1357
atgaaattta	acttactcaa	agttacacag	aactcctggc	tgatctcttt	tacttaattc	1417
ttctaaaatc	taaagtaaga	ataattttga	aattttattg	gtaatgctta	aataacaaaa	1477
gttactttta	tatgttacat	gtttgaagtt	tcattttatac	ttaatatcta	atttaacgct	1537
tgtagaaata	ttattcagta	tttttttttc	ttttctaaaa	tgacttggtc	agggctcctt	1597
ttgggatttt	aagtgagaaa	aactctatct	cataatattg	ttaccatgct	catccaataa	1657
cacttcatag	atggttagcc	atctctaaaa	ttaaattttg	tagttgagga	gcattgaaaa	1717
tatcacctat	atttggattt	ctgtttaaca	cttcaaaatg	cattttatca	taatttaaaa	1777
tttgatttaa	atacaaacct	ctttatagaa	aacctgataa	atgtagtttt	atttgtttgc	1837
ttgctttttt	ttgcagggag	ggatattccc	ttgataccac	tctgtcttta	tgactcatcg	1897
gttattttata	aatgttatca	tcgtttttct	gatttttaaa	taagtttcag	aattaataca	1957
tttaactgaa	taaatacaacc	aataatgtca	aactt			1992

&lt;210&gt; 8

&lt;211&gt; 1898

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; 5'UTR

&lt;222&gt; 1..95

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 96..182

&lt;220&gt;

&lt;221&gt; 3'UTR

&lt;222&gt; 183..1898

&lt;220&gt;

&lt;221&gt; polyA\_signal

&lt;222&gt; 1872..1877

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 132

&lt;223&gt; 8-130-143 : polymorphic base A or G

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 132

&lt;223&gt; 8-126-286 : polymorphic base A or G

&lt;400&gt; 8

ttgcaggtcc acattcggtc tagaggtgc ctgcggtcct tagaggcctc ataggcaagg 60  
 aactcatagt ttcttgccat gaagactttt ccaac atg gct gct tac ttc ttc 113

Met Ala Ala Tyr Phe Phe

1

5

aag cct gta agg aga tcc rgt ctc agt tgg cca ggt ctt gtg aaa ggg 161

Lys Pro Val Arg Ser Xaa Leu Ser Trp Pro Gly Leu Val Lys Gly

10

15

20

ttt tgt gca tct cta aga tga tcatcttggt gtggggctcct gatcttccca 212

Phe Cys Ala Ser Leu Arg \*

25

gtgacttagc agcacagtaa tttggctgaa atgttgatga gttcagacaa gaagaaggtg 272

atgatgacaa gagagcgtca cggggaaatt ctgccctgag gaaggaaatag tttgaggctg 332

gagagatgtg ctacccccgc cactctttct caaatggaaa gacaacctgt aggcacagaa 392

atcattgctc tgcttacact gaaatctcct gctgactttg agagctgagc attggatttt 452

gatgggagag tgaacttag tgaagaaatg atttgaacct ttgttgatct aatacttctg 512

tgatacagtt atctaaacag aatgagtcta aagtcagtat aaagacctgt gtctctccat 572

gttttgctgc ttataattgc agaggaacat tttgagatag aaaattttat cctcttttca 632

agtgactttt tataaaacat aaccttttaa gagaggtcat tcttttgaaa aattgtctga 692

atacaaaaaa gtaactgaac agcaagggat tgtagccata atgttcttga ccactttaga 752

ttgacaaata gtatctcaaa attgatcagg ttctggagaa tctattgagg ttgtatggag 812

taaccaagtg atggaagagg ttggaagtga ataattttat gtggtattat gtgatttttt 872

ttggtggggg tagtgtctgg agctgagcat gatggagcct agatggggca gccttctaatt 932

gctcccagta tttcgcaaat catatgtcac gtcaaaattg taatgacaaa gctcacagca 992

gctgcctacg ccaggtttgg tgccagctct aattagagca ttgatcattt ctctcccctt 1052

catgaatgca catgagtgc ctttttcata tataagggca aaggacattt caaagtttgg 1112

atggaatctt tctacttag aagtaataaa aaacaatact ttgaaatctt accaaatggt 1172

aaacaaatag caacaatttt tccatttgaa gatgtgtgtc ttaagggtctg ttggatcttt 1232

caaattttcta ttttgacat ggaaactaca gatgaaattt aacttactca aagttacaca 1292

gaactcctgg ctgatctctt ttacttaatt ctcttaaaat cttaaagtaag aataattttg 1352

aaattttatt ggtaatgctt aaataacaaa agttacttta atatgttaca tgtttgaagt 1412

ttcatttata cttaatatct aatttaacgc ttgtgaaaat attattcagt attttttttt 1472

cttttctaaa atgacttggt cagggctcct tttgggattt taagtgaaga aaactctatt 1532

tcataatatt gttaccatgc tcatccaata acacttcata gatggttagc catctctaaa 1592

attaaatttt gtagttagg agcattgaaa atatcaccta ttttggatt tctgtttaac 1652

acttcaaaat gcattttatc ataattttaa atttgattta aatacaaac tctttataga 1712

aaacctgata aatgtagttt tatttgtttg ctgtcttttt tttgcaggga gggatattcc 1772

cttgatacca ctctgtcttt atgactcatc ggttatttat aaatgttatc atogtttttc 1832

tgatttttaa ataagtttca gaattaatac atttaactga ataaatcaac caataatgtc 1892

aaactt 1898

&lt;210&gt; 9

&lt;211&gt; 1783

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; 5'UTR

&lt;222&gt; 1..95

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 96..167

&lt;220&gt;

&lt;221&gt; 3'UTR

&lt;222&gt; 168..1783

&lt;220&gt;

&lt;221&gt; polyA\_signal

&lt;222&gt; 1757..1762

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 132

&lt;223&gt; 8-130-143 : polymorphic base A or G

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 132

&lt;223&gt; 8-126-286 : polymorphic base A or G

&lt;400&gt; 9

ttgcaggctcc	acattcggttc	tagaggctgc	ctgcggctcct	tagaggcctc	ataggcaagg	60
aactcatagt	ttcttgccat	gaagactttt	ccaac	atg gct gct tac ttc ttc		113

	Met	Ala	Ala	Tyr	Phe	Phe
1				5		

aag cct gta agg aga tcc rgt ctc agt tgg cca ggt ctt gtg aaa gga	161
Lys Pro Val Arg Arg Ser Xaa Leu Ser Trp Pro Gly Leu Val Lys Gly	

10

15

20

aaa taa cgcatttata aaagccagct cagatgtttac ttccaaccaa caaatttctc	217
Lys *	

ggacacacta	agtaatagag	ttaatcacac	cgtcctgaaa	gacaacctgt	aggcacagaa	277
atcattgctc	tgcttacact	gaaatctcct	gctgactttg	agagctgagc	attggatttt	337
gatgggagag	tgaaacttag	tgaagaaatg	atttgaacct	ttgttgatct	aatacttctg	397
tgatacagtt	atctaaacag	aatgagtcta	aagtcagtat	aaagacctgt	gtctctccat	457
gttttgctgc	ttataattgc	agaggaacat	tttgagatag	aaaattttat	ccctctttca	517
agtgactttt	tataaaacat	aaccttttaa	gagaggtcat	tcttttgaaa	aattgtctga	577
atacaaaaaa	gtaactgaac	agcaagggat	tgtagccata	atgttcttga	ccactttaga	637
ttgacaaata	gtatctcaaa	attgatcagg	ttctggagaa	tctattgagg	ttgtatggag	697
taaccaagtg	atggaagagg	ttggaagtga	ataattttat	gtggtattat	gtgatttttt	757
ttggtggggg	tagtgtctgg	agctgagcat	gatggagcct	agatggggca	gccttcta	817
gctcccagta	tttcgcaaat	catatgtcac	gtcaaaattg	taatgacaaa	gctcacagca	877
gctgcctacg	ccagggttgg	tgccagctct	aattagagca	ttgatcattt	ctctcccctt	937
catgaatgca	catgagtgc	cattttcata	tataagggca	aaggacattt	caaagtttgg	997
atggaatcct	tcctacttag	aagtaataaa	aaacaatact	ttgaaatcct	accaaatggt	1057
aaacaaatag	caacaatttt	tocatttgaa	gatgtgtgtc	ttaagggtctg	ttggatcttt	1117
caaatttcta	ttttggacat	ggaaactaca	gatgaaattt	aacttactca	aagttacaca	1177
gaactcctgg	ctgatctctt	ttacttaatt	cttctaaaat	ctaaagtaag	aataattttg	1237
aaattttatt	ggtaatgctt	aaataacaaa	agttacttta	atatgttaca	tgtttgaagt	1297
ttcatttata	cttaatatct	aatttaacgc	ttgtgaaaat	attattcagt	attttttttt	1357
cttttctaaa	atgacttggt	cagggtcct	tttgggattt	taagtgagaa	aaactctatt	1417
tcataatatt	gttaccatgc	tcatacaata	acacttcata	gatgggttagc	catctctaaa	1477
attaaatttt	gtagttgagg	agcattgaaa	atatcaccta	tatttggtatt	tctgtttaac	1537
acttcaaaat	gcattttatc	ataatttaaa	atttgattta	aatacaaaac	tctttataga	1597
aaacctgata	aatgtagttt	tatttgtttg	cttgcttttt	tttgcaggga	gggatattcc	1657
cttgatacca	ctctgtcttt	atgactcctc	gggtattttat	aaatgttatc	atcgtttttc	1717
tgatttttaa	ataagtttca	gaattaatac	atttaactga	ataaatcaac	caataatgtc	1777
aaactt						1783

&lt;210&gt; 10

&lt;211&gt; 2309

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; 5'UTR

&lt;222&gt; 1..95

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 96..311

&lt;220&gt;

&lt;221&gt; 3'UTR

&lt;222&gt; 312..2309

&lt;220&gt;

&lt;221&gt; polyA\_signal

&lt;222&gt; 2283..2288

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 132

&lt;223&gt; 8-130-143 : polymorphic base A or G

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 132

&lt;223&gt; 8-126-286 : polymorphic base A or G

&lt;400&gt; 10

ttgcaggctcc acattcgctc tagaggctgc ctgcggctcct tagaggcctc ataggcaagg 60  
 aactcatagt ttcttgccat gaagactttt ccaac atg gct gct tac ttc ttc 113

Met Ala Ala Tyr Phe Phe  
 1 5

aag cct gta agg aga tcc rgt ctc agt tgg cca ggt ctt gtg aaa ggt 161  
 Lys Pro Val Arg Arg Ser Xaa Leu Ser Trp Pro Gly Leu Val Lys Gly  
 10 15 20

aac ctc atc acc gga gtc aag tgt cat cac ctt tcc cat gtt cca ttg 209  
 Asn Leu Ile Thr Gly Val Lys Cys His His Leu Ser His Val Pro Leu  
 25 30 35

gtt aga agt ccc agc tcc tgc cca cat tca agg gaa aat aac gca ttt 257  
 Val Arg Ser Pro Ser Ser Cys Pro His Ser Arg Glu Asn Asn Ala Phe  
 40 45 50

ata aaa gcc agc tca gat gtt act tcc aac caa att tcc tgg aca 305  
 Ile Lys Ala Ser Ser Asp Val Thr Ser Asn Gln Gln Ile Ser Trp Thr  
 55 60 65 70

cac taa gtaatagagt taatcacacc gtccctgaaca ttcaccctgg gagctctgtg 361  
 His \*

ctgccatgca aatgtcacat tgcactgagg ccgccatgct gtaaaaaacc ccaagctacc 421

tacctgaaaa gctggagaga atcagcaccc atcagctttt ccagcatttt gggctcctccc 481

agctaaatca tcggcattgt gaagcagaga tgagacctcc tgccatgccc ttcccaaacc 541

ctggaccaac agaatcatca tatacactag gttttgtgca tctctaagat gatcatcttg 601

gtgtggggtc ctgatcttcc cagtgaactta gcagcacagt aatttggctg aaatgttgat 661

gagttcagac aagaagaagg tgatgatgac aagagagcgt cacggggaaa ttctgccttg 721

aggaaggaat agtttgaggc tggagagatg tgctaccccc gccactcttt ctcaaatgga 781

aagacaacct gtaggcacag aaatcattgc tctgcttaca ctgaaatctc ctgctgactt 841

tgagagctga gcattggatt ttgatgggag agtgaaactt agtgaagaaa tgatttgaac 901

ctttgttgat ctaatacttc tgtgatacag ttatctaaac agaattgagtc taaagtcagt 961

ataaagacct gtgtctctcc atgttttgct gcttataatt gcagaggaac attttgagat 1021

agaaaatttt atccctcttt caagtgaactt tttataaaac ataacctttt aagagaggtc 1081

attcttttga aaaattgtct gaatacaaaa aagtaactga acagcaaggg attgtagcca 1141

taatgttctt gaccacttta gattgacaaa tagtatctca aaattgatca ggttctggag 1201

aatctattga ggttgatgag agtaaccaag tgatggaaga ggttggaagt gaataatttt 1261

atgtggtatt atgtgatttt ttttgggtggg ggtagtgtct ggagctgagc atgatggagc 1321

ctagatgggg cagccttcta atgctccag tatttcgcaa atcatatgtc acgtcaaaat 1381

tgtaatgaca aagctcacag cagctgccta cgccagggtt ggtgccagct ctaattagag 1441

cattgatcat ttctctcccc ttcatgaatg cacatgagtg cacattttca tatataaggg 1501

caaaggacat ttcaaagttt ggatggaatc tttcctaactt agaagtaata aaaaacaata 1561

ctttgaaatc ttaccaaattg ttaaacaaat agcaacaatt tttccatttg aagatgtgtg 1621

tcttaaggtc tgttggatct ttcaaatctt tattttggac atggaaacta cagatgaaat 1681

```

ttaacttact caaagttaca cagaactcct ggctgatctc ttttacttaa ttcttctaaa 1741
atctaaagta agaataattt tgaaatttta ttggtaatgc ttaaataaca aaagttactt 1801
taatatgtta catgtttgaa gtttcattta tacttaatat ctaatttaac gcttgtgaaa 1861
atattattca gtattttttt ttcttttcta aaatgacttg ttcagggtc cttttgggat 1921
tttaagttag aaaaactcta tttcataata ttgttaccat gtcacatcaa taacacttca 1981
tagatggtta gccatctcta aaattaaatt ttgtagttga ggagcattga aaatatcacc 2041
tatatttgga tttctgttta acacttcaaa atgcatttta tcataattta aaatttgatt 2101
taaatacaaaa cctctttata gaaaacctga taaatgtagt tttatttggt tgcttgcttt 2161
tttttgcagg gagggatatt cccttgatac cactctgtct ttatgactca tcggttattt 2221
ataaatgtta tcatcgttt tctgattttt aaataagttt cagaattaat acatttaact 2281
gaataaatca accaataatg tcaaactt 2309

```

<210> 11  
 <211> 2075  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> 5'UTR  
 <222> 1..95

<220>  
 <221> CDS  
 <222> 96..311

<220>  
 <221> 3'UTR  
 <222> 312..2075

<220>  
 <221> polyA\_signal  
 <222> 2049..2054

<220>  
 <221> allele  
 <222> 132  
 <223> 8-130-143 : polymorphic base A or G

<220>  
 <221> allele  
 <222> 132  
 <223> 8-126-286 : polymorphic base A or G

```

<400> 11
ttgcagggtcc acattcggtc tagagggtgc ctgcgggtcct tagaggcctc ataggcaagg 60
aactcatagt ttcttgccat gaagactttt ccaac atg gct gct tac ttc ttc 113
                               Met Ala Ala Tyr Phe Phe
                               1         5
aag cct gta agg aga tcc rgt ctc agt tgg cca ggt ctt gtg aaa ggt 161
Lys Pro Val Arg Arg Ser Xaa Leu Ser Trp Pro Gly Leu Val Lys Gly
                               10        15        20
aac ctc atc acc gga gtc aag tgt cat cac ctt tcc cat gtt cca ttg 209
Asn Leu Ile Thr Gly Val Lys Cys His His Leu Ser His Val Pro Leu
                               25        30        35
gtt aga agt ccc agc tcc tgc cca cat tca agg gaa aat aac gca ttt 257
Val Arg Ser Pro Ser Ser Cys Pro His Ser Arg Glu Asn Asn Ala Phe
                               40        45        50
ata aaa gcc agc tca gat gtt act tcc aac caa caa att tcc tgg aca 305
Ile Lys Ala Ser Ser Asp Val Thr Ser Asn Gln Gln Ile Ser Trp Thr
                               55        60        65        70
cac taa gtaatagagt taatcacacc gtcttggttt tgtgcatctc taagatgatc 361

```

His \*

```

atcttggtgt ggggtcctga tcttcccagt gacttagcag cacagtaatt tggctgaaat 421
gttgatgagt tcagacaaga agaagggtgat gatgacaaga gagcgtcacg gggaaattct 481
gccctgagga aggaatagtt tgaggctgga gagatgtgct acccccgcca ctctttctca 541
aatggaaaga caacctgtag gcacagaaat cattgtctctg cttacactga aatctcctgc 601
tgactttgag agctgagcat tggattttga tgggagagtg aaacttagtg aagaaatgat 661
ttgaaccttt gttgatctaa tacttctgtg atacagttat ctaaacagaa tgagtctaaa 721
gtcagtataa agacctgtgt ctctccatgt tttgctgctt ataattgcag aggaacattt 781
tgagatagaa aattttatoc ctctttcaag tgacttttta taaaacataa ctttttaaga 841
gagggtcattc ttttgaaaaa ttgtctgaat acaaaaaagt aactgaacag caagggattg 901
tagccataat gttcttgacc acttttagatt gacaaatagt atctcaaaat tgatcaggtt 961
ctggagaatc tattgaggtt gtatggagta accaagtgat ggaagaggtt ggaagtgaat 1021
aattttatgt ggtattatgt gatttttttt ggtgggggta gtgtctggag ctgagcatga 1081
tggagcctag atggggcagc cttctaagtc tcccagttat tcgcaaatca tatgtcacgt 1141
caaaattgta atgacaagc tcacagcagc tgcctacgcc aggtttggtg ccagctctaa 1201
ttagagcatt gatcatctt ctccccttca tgaatgcaca tgagtgcaca ttttcatata 1261
taagggcaaa ggacatttca aagtttggat ggaatcttct ctacttagaa gtaataaaaa 1321
acaatacttt gaaatcttac caaatgttaa acaaatagca acaatttttc catttgaaga 1381
tgtgtgtctt aaggctctgtt ggatctttca aatttctatt ttggacatgg aaactacaga 1441
tgaaatttaa cttactcaaa gttacacaga actcctggct gatctctttt acttaattct 1501
tctaaaatct aaagtaagaa taattttgaa atttttattgg taatgcttaa ataacaaaag 1561
ttactttaat atgtttacatg tttgaagttt catttatact taatatctaa tttaacgctt 1621
gtgaaaatat tattcagtat ttttttttct tttctaaaat gacttggtca gggctccttt 1681
tgggatttta agtgagaaaa actctatttc ataataattgt taccatgctc atccaataac 1741
acttcataga tggtttagcca tctctaaaaa taaattttgt agttgaggag cattgaaaat 1801
atcacctata tttggatttc tgtttaacac ttcaaaatgc attttatcat aatttataat 1861
ttgattttaa tacaacctc tttatagaaa acctgataaa tgtagtttta tttgtttgct 1921
tgcttttttt tgcagggagg gatattccct tgataccact ctgtctttat gactcatcgg 1981
ttatttataa atgttatcat cgtttttctg attttttaaa aagtttcaga attaatacat 2041
ttaactgaat aaatcaacca ataatgtcaa actt 2075

```

&lt;210&gt; 12

&lt;211&gt; 1923

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; 5'UTR

&lt;222&gt; 1..95

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 96..212

&lt;220&gt;

&lt;221&gt; 3'UTR

&lt;222&gt; 213..1923

&lt;220&gt;

&lt;221&gt; polyA\_signal

&lt;222&gt; 1897..1902

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 132

&lt;223&gt; 8-130-143 : polymorphic base A or G

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 132

&lt;223&gt; 8-126-286 : polymorphic base A or G

&lt;400&gt; 12

ttgcagggtcc acattcggtc tagagggtgc ctgcgggtcct tagaggcctc ataggcaagg 60  
 aactcatagt ttcttgccat gaagactttt ccaac atg gct gct tac ttc ttc 113

Met Ala Ala Tyr Phe Phe

1

5

aag cct gta agg aga tcc rgt ctg agt tgg cca ggt ctt gtg aaa gga 161  
 Lys Pro Val Arg Arg Ser Xaa Leu Ser Trp Pro Gly Leu Val Lys Gly

10

15

20

aca ttc acc ctg gga gct ctg tgc tgc cat gca aat gtc aca ttg cac 209

Thr Phe Thr Leu Gly Ala Leu Cys Cys His Ala Asn Val Thr Leu His

25

30

35

tga ggccgccatg ctgtataaaaa ccccaagcta cctacctgaa aagctggaga 262

\*

gaatcagcac ccattcagctt ttccagcatt ttgggtcctc ccagctaaat catcggcatt 322  
 gtgaagcaga gatgagacct cctgccatgc cctttccaaa tcttgacca acagaatcat 382  
 catatacact agaaagacaa cctgtaggca cagaaatcat tgctctgctt acactgaaat 442  
 ctctctgctga ctttgagagc tgagcattgg attttgatgg gagagtgaat cttagtgaag 502  
 aaatgatttg aacctttggt gatctaatac ttctgtgata cagttatcta aacagaatga 562  
 gtctaaagtc agtataaaga cctgtgtctc tccatgtttt gctgcttata attgcagagg 622  
 aacattttga gatagaaaaa ttatccctc ttccaagtga ctttttataa aacataacct 682  
 ttttaagagag gtcattcttt tgaaaaattg tctgaataca aaaaagtaac tgaacagcaa 742  
 gggattgtag ccataatggt cttgaccact ttagattgac aaatagtatc tcaaaattga 802  
 tcagggttctg gagaatctat tgaggttgta tggagtaacc aagtgatgga agagggttga 862  
 agtgaataat tttatgtggt attatgtgat tttttttggt gggggtagtg tctggagctg 922  
 agcatgatgg agcctagatg gggcagcctt ctaatgctcc cagtatttcc caaatcatat 982  
 gtcacgtcaa aattgtaattg acaaaagctca cagcagctgc ctacgccagg tttggtgcc 1042  
 gctctaatta gagcattgat cttttctctc cccttcatga atgcacatga gtgcacattt 1102  
 tcatatataa gggcaaaagg catttcaaag ttgggatgga atctttccta cttagaagta 1162  
 ataaaaaaca atactttgaa atcttaccac atgttaaaaca aatagcaaca atttttccat 1222  
 ttgaagatgt gtgtcttaag gtctgttggga tctttcaaat ttctattttg gacatggaaa 1282  
 ctacagatga aatttaactt actcaaagtt acacagaact cctggctgat ctcttttact 1342  
 taattcttct aaaatctaaa gtaagaataa ttttgaaatt ttatttgtaa tgcttaataa 1402  
 acaaaagtta ctttaatatg ttacatgttt gaagtttcat ttatacttaa tatctaattt 1462  
 aacgcttggtg aaaatattat tcagtatttt tttttctttt ctaaaatgac ttgttcaggg 1522  
 ctcccttttg gatttttaag gagaaaaact ctatttcata atattgttac catgctcatc 1582  
 caataacact tcatagatgg ttagccatct ctaaaattaa attttgtagt tgaggagcat 1642  
 tgaaatatc acctatattt ggatttctgt ttaacacttc aaaatgcatt ttatcataat 1702  
 ttaaaatttg atttaaatat aaacctcttt atagaaaacc tgataaatgt agttttatatt 1762  
 gtttgcttgc ttttttttgc agggagggat attcccttga taccactctg tctttatgac 1822  
 tcatcggtta tttataaatg ttatcatcgt tttctgatt tttaataag tttcagaatt 1882  
 aatacattta actgaataaa tcaaccaata atgtcaaaact t 1923

&lt;210&gt; 13

&lt;211&gt; 2308

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; 5'UTR

&lt;222&gt; 1..95

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 96..167

&lt;220&gt;

&lt;221&gt; 3'UTR

&lt;222&gt; 168..2308

&lt;220&gt;

<221> polyA\_signal  
<222> 2282..2287

<220>  
<221> allele  
<222> 132  
<223> 8-130-143 : polymorphic base A or G

<220>  
<221> allele  
<222> 132  
<223> 8-126-286 : polymorphic base A or G

<400> 13  
ttgcagggtcc acatttcgttc tagagggtgc ctgcgggtcct tagaggcctc ataggcaagg 60  
aactcatagt ttcttgccat gaagactttt ccaac atg gct gct tac ttc ttc 113  
Met Ala Ala Tyr Phe Phe  
1 5  
aag cct gta agg aga tcc rgt ctc agt tgg cca ggt ctt gtg aaa gga 161  
Lys Pro Val Arg Arg Ser Xaa Leu Ser Trp Pro Gly Leu Val Lys Gly  
10 15 20  
aaa taa cgcatttata aaagccagct cagatgttac ttccaaccaa caaatttcct 217  
Lys \*  
ggacacacta agtaatatagag ttaatcacac cgtcctaaac aatacgaaca gaaacaaagc 277  
tatgtacttg ttttgggaag caacctcaaa tggaaaagatg ggggctgtgt tctaaaagga 337  
acattcaccc tgggagctct gtgctgccat gcaaagtca cattgcactg aggccgccat 397  
gctgtaaaaa accccaagct acctacctga aaagctggag agaatcagca cccatcagct 457  
tttcagcat tttgggtcct cccagctaaa tcatcgcat tgtgaagcag agatgagacc 517  
tctgccatg ccctttccaa atcctggacc aacagaatca tcatatacac taggttttgt 577  
gcactctctaa gatgatcatc ttggtgtggg gtcctgatct tccagtgac ttagcagcac 637  
agtaatttgg ctgaaatgtt gatgagttca gacaagaaga aggtgatgat gacaagagag 697  
cgtcacgggg aaattctgcc ctgaggaagg aatagtttga ggctggagag atgtgctacc 757  
cccgccactc tttctcaaat ggaaagacaa cctgtaggca cagaaatcat tgctctgctt 817  
acactgaaat ctctgctga ctttgagagc tgagcattgg attttgatgg gagagtgaag 877  
cttagtgaag aaatgatttg aacctttgtt gatctaatac ttctgtgata cagttatcta 937  
aacagaatga gtctaaagtc agtataaaga cctgtgtctc tccatgtttt gctgcttata 997  
attgcagagg aacattttga gatagaaaat tttatccctc tttcaagtga ctttttataa 1057  
aacataacct ttaagagag gtcattcttt tgaaaaattg tctgaatata aaaaagtaac 1117  
tgaacagcaa gggattgtag ccataatgtt cttgaccact ttagattgac aaatagtatc 1177  
tcaaaaattga tcaggttctg gagaatctat tgaggtttga tggagtaacc aagtgatgga 1237  
agaggttga agtgaataat tttatgtggt attatgtgat ttttttggg gggggtagt 1297  
tctggagctg agcatgatgg agcctagatg gggcagcctt ctaatgtctc cagtatttcg 1357  
caaatcatat gtcacgtcaa aattgtaatg acaaagctca cagcagctgc ctacgccagg 1417  
tttggtgcca gctctaatta gagcattgat ctttctctc cccttcatga atgcacatga 1477  
gtgcacattt tcatatataa gggcaaaagg catttcaaag tttggatgga atctttccta 1537  
cttagaagta ataaaaaaca atactttgaa atcttaacca atgttaaaca aatagcaaca 1597  
atttttccat ttgaagatgt gtgtcttaag gtctgttggg tctttcaaat ttctattttg 1657  
gacatggaaa ctacagatga aatttaactt actcaaagtt acacagaact cctggctgat 1717  
ctcttttaact taattcttct aaaaatctaaa gtaagaataa ttttgaaatt ttattggtaa 1777  
tgcttaataa acaaaagtta cttaatatg ttacatgttt gaagtttcat ttatacttaa 1837  
tatctaattt aacgcttggtg aaaaatattat tcagtatttt tttttctttt ctaaaatgac 1897  
ttgttcaggg ctcttttggg gattttaagt gagaaaaact ctatttcata atattgttac 1957  
catgctcatc caataacact tcatagatgg ttagccatct ctaaaattaa attttgtagt 2017  
tgaggagcat tgaaaatatc acctatattt ggatttctgt ttaacacttc aaaatgcatt 2077  
ttatcataat ttaaaatttg atttaaatat aaacctcttt atagaaaacc tgataaatgt 2137  
agttttattt gtttgcttgc ttttttttgc agggagggat attcccttga taccactctg 2197  
tctttatgac tcatcggtta tttataaatg ttatcatcgt ttttctgatt tttaaataag 2257  
tttcagaatt aatacattta actgaataaa tcaaccaata atgtcaaact t 2308

<210> 14  
<211> 2706





```

ataatgttct tgaccacttt agattgacaa atagtatctc aaaattgac aggttctgga 1597
gaatctattg aggttgatg gagtaaccaa gtgatggaag .aggttggaag tgaataattt 1657
tatgtggtat tatgtgattt tttttggtgg gggtagtgct tggagctgag catgatggag 1717
cctagatggg gcagccttct aatgctccca gtatttcgca aatcatatgt cacgtcaaaa 1777
ttgtaatgac aaagctcaca gcagctgcct acgccaggtt tggtgccagc tctaattaga 1837
gcattgatca tttctctccc cttcatgaat gcacatgagt gcacattttc atatataagg 1897
gcaaaggaca tttcaaagtt tggatggaat ctttccact tagaagtaat aaaaaacaat 1957
actttgaaat cttaccaaat gttaacaaa tagcaacaat tttccattt gaagatgtgt 2017
gtccttaagg ctgttggtatc tttcaaattt ctattttgga catggaaact acagatgaaa 2077
tttaacttac tcaaagttac acagaactcc tggctgatct cttttactta attcttctaa 2137
aatctaaagt aagaataatt ttgaaatttt atttgtaatg cttaaataac aaaagttact 2197
ttaatatggt acatgtttga agtttctatt atacttaata tctaatttaa cgcttgtgaa 2257
aatattattc agtatttttt tttcttttct aaaatgactt gttcagggtc ccttttgga 2317
ttttaagtga gaaaaactct atttcataat attggtacca tgctcatcca ataacacttc 2377
atagatgggt agccatctct aaaattaaat tttgtagttg aggagcattg aaaatatcac 2437
ctatatttgg atttctgttt aacacttcaa aatgcatttt atcataattt aaaatttgat 2497
ttaaatacaa acctctttat agaaaacctg ataaatgtag ttttatttgt ttgcttgctt 2557
ttttttgcag ggagggatat tcccttgata ccactctgtc tttatgactc atcggttatt 2617
tataaatggt atcatcggtt ttctgatttt taaataagtt tcagaattaa tacatttaac 2677
tgaataaatc aaccaataat gtcaaactt 2706

```

&lt;210&gt; 15

&lt;211&gt; 2612

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; 5'UTR

&lt;222&gt; 1..95

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 96..212

&lt;220&gt;

&lt;221&gt; 3'UTR

&lt;222&gt; 213..2612

&lt;220&gt;

&lt;221&gt; polyA\_signal

&lt;222&gt; 2586..2591

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 132

&lt;223&gt; 8-130-143 : polymorphic base A or G

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 132

&lt;223&gt; 8-126-286 : polymorphic base A or G

&lt;400&gt; 15

```

ttgcaggctc acattcggtc tagaggctgc ctgcggctcct tagaggcctc ataggcaagg 60
aactcatagt ttcttgccat gaagactttt ccaac atg.gct gct tac ttc ttc 113
                               Met Ala Ala Tyr Phe Phe
                               1         5
aag cct gta agg aga tcc rgt ctc agt tgg cca ggt ctt gtg aaa gga 161
Lys Pro Val Arg Arg Ser Xaa Leu Ser Trp Pro Gly Leu Val Lys Gly
                               10        20
aca ttc acc ctg gga gct ctg tgc tgc cat gca aat gtc aca ttg cac 209

```

Thr Phe Thr Leu Gly Ala Leu Cys Cys His Ala Asn Val Thr Leu His  
 25 30 35  
 tga ggcgcgcatg ctgtataaaaa ccccaagcta cctacctgaa aagctggaga 262  
 \*  
 gaatcagcac ccacagcgtt ttccagcatt ttgggtcctc ccagctaaat catcggcatt 322  
 gtgaagcaga gatgagacct cctgcatgac cctttccaaa tcctggacca acagaatcat 382  
 catatacact agtaagcttt tactttaagc cactaatctg ggggtgtagt cttttgcagc 442  
 aaaagatcat cagaacactg ccacgcagga gctgataaac ttgctcactg ttatgtatgt 502  
 atccaatgac tactcacaca cacattactt gcttcactct actttctggc tttccattca 562  
 gcaaatttta agaaatgaaa atgttattcg tgggaatta actcaagatc aatattgttg 622  
 tagttaatct ctttgttaaat ggtataaatat taaattattg tgagtcattt ttatttaag 682  
 aagtttggtt ctttaaatga aagaagtttg ggtcttttgt ggatactctg catatagaca 742  
 ttctagacat gaaagaatgt tgaagacaaa gaattgtgaa gaaatataga catgaaagac 802  
 tgttgaagtt tgggttttttg tggatactct gcatatagac attctagact ggagtgatag 862  
 atcttgttca ggggttttgt catctctaag atgatcatct tgggtgtggg tcctgatctt 922  
 cccagtgact tagcagcaca gtaatttggc tgaaatgttg atgagttcag acaagaagaa 982  
 ggtgatgatg acaagagagc gtcacgggga aattctgccc tgaggaagga atagtttgag 1042  
 gctggagaga tgtgctaccc ccgccactct ttctcaaag gaaagacaac ctgtaggcac 1102  
 agaaatcatt gctctgctta cactgaaatc tcctgctgac ttgagagct gagcattgga 1162  
 ttttgatggg agagtgaac ttagtgaaga aatgatttga acccttgttg atctaatact 1222  
 tctgtgatac agttatctaa acagaatgag tctaaagtca gtataagac ctgtgtctct 1282  
 ccattgtttg ctgcttataa ttgcagagga acattttgag atagaaaatt ttatccctct 1342  
 ttcaagtga tttttataaa acataacctt ttaagagagg tcattctttt gaaaaattgt 1402  
 ctgaatacaa aaaagtaact gaacagcaag ggattgtagc cataatgttc ttgaccactt 1462  
 tagattgaca aatagtatct caaaattgat caggttctgg agaattctatt gaggttgat 1522  
 ggagtaacca agtgatggaa gaggttgga gtgaataatt ttatgtggtt ttatgtgatt 1582  
 ttttttggtg ggggtagtgt ctggagctga gcatgatgga gcctagatgg ggcagccttc 1642  
 taatgctccc agtatttctgc aatcatatg tcacgtcaaa attgtaatga caaagctcac 1702  
 agcagctgcc tacgccaggt ttggtgccag ctctaattag agcattgatc atttctctcc 1762  
 ccttcagaa tgcacatgag tgcacatttt catatataag ggcaaaggac atttcaaagt 1822  
 ttggatggaa tctttcctac ttagaagtaa taaaaaaca tactttgaaa tcttaccaa 1882  
 tgttaacaa atagcaacaa tttttccatt tgaagatgtg tgtcttaagg tctgttgat 1942  
 ctttcaaatt tctatttttg acatggaaac tacagatgaa atttaactta ctcaaagtta 2002  
 cacagaactc ctggctgatc tcttttactt aattcttcta aaatctaaag taagaataat 2062  
 ttgaaattt tattggtaat gottaaataa caaaagttac tttaatatgt tacatgtttg 2122  
 aagtttcatt tatacttaat atctaattta acgcttgtga aaatattatt cagtattttt 2182  
 ttttctttc taaaatgact tgttcagggc tccttttggg attttaagtg agaaaaactc 2242  
 tatttcataa tattgttacc atgctcatcc aataacactt catagatggt tagccatctc 2302  
 taaaattaa ttttgtagtt gaggagcatt gaaaaatca cctatatattg gatttctgtt 2362  
 taacacttca aaatgcattt tatcataatt taaaatttga tttaaataca aacctcttta 2422  
 tagaaaacct gataaatgta gttttatttg tttgctgtct tttttttgca gggagggata 2482  
 ttcccttgat accactctgt ctttatgact catcggttat ttataaatgt tatcatcgtt 2542  
 tttctgattt ttaaataagt ttcagaatta atacatttaa ctgaataaat caaccaataa 2602  
 tgtcaaaactt 2612

<210> 16  
 <211> 434  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> 5'UTR  
 <222> 1..370

<220>  
 <221> CDS  
 <222> 371..436

<220>  
 <221> 3'UTR  
 <222> 437..434

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 407

&lt;223&gt; 8-130-143 : polymorphic base A or G

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 407

&lt;223&gt; 8-126-286 : polymorphic base A or G

&lt;400&gt; 16

tcccagtg	ggg attcaaccac	tcatagactg	cttttggtca	aactgcgttc	aaataaggca	60
aacgccc	gagc tggaaccaat	ctcactgttt	ctgtacctcg	cttctaattc	cttttcagaa	120
aacaaaa	act cctaagagaa	tgaatagaaa	ttttcttaag	gaattataaa	gaaccatctc	180
agaatttt	ac tctgttactg	caaaatctca	tgcaaagata	accaagtacg	agaattaac	240
tgaaatt	cat cccaagatga	ggcatcttgg	ctgagttgca	ggtccacatt	cgttctagag	300
gctgcct	gcg gtccttagag	gcctcatagg	caaggaactc	atagtttctt	gccatgaaga	360
cttttcca	ac atg gct gct	tac ttc ttc	aag cct gta	agg aga	tcc rgt	409

Met Ala Ala Tyr Phe Phe Lys Pro Val Arg Arg Ser Xaa

1

5

10

ctc agt tgg cca ggt ctt gtg aaa g

434

Leu Ser Trp Pro Gly Leu Val Lys

15

20

&lt;210&gt; 17

&lt;211&gt; 760

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; 5'UTR

&lt;222&gt; 1..696

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; 697..762

&lt;220&gt;

&lt;221&gt; 3'UTR

&lt;222&gt; 763..760

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 733

&lt;223&gt; 8-130-143 : polymorphic base A or G

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 733

&lt;223&gt; 8-126-286 : polymorphic base A or G

&lt;400&gt; 17

tcccagtg	ggg attcaaccac	tcatagactg	cttttggtca	aactgcgttc	aaataaggca	60
aacgccc	gagc tggaaccaat	ctcactgttt	ctgtacctcg	cttctaattc	ctgtacgtca	120
ctttacct	ttt tttgtctata	aatttattct	gaccacaagg	cacccctgga	gcctctgtga	180
gtctgct	gtg gttctgggga	ctgccggatt	cgcgaatcgt	tcattttctca	gttaaactcc	240
tttaaatta	aa ttcagctgaa	gtttttcttt	tatcaatgaa	aagactagac	tggttagacc	300
ttccagc	ectg cacttttctc	ccatgctgga	tgtttcctgc	ccttgaatat	cagactccaa	360
gtttctc	agtt tttgggactc	agactggctt	ccttgcctct	cagcttgag	atgccctatt	420
gtgggg	gtctt gtttgcgttt	tcagaaaaca	aaaactccta	agagaatgaa	tagaaatttt	480

WO 01/40493

PCT/IB00/01507

cttaaggaat tataaagaac catctcagaa ttttactctg ttactgcaaa atctcatgca 540  
 aagataacca agtacgaaga attaactgaa attcatccca agatgaggca tcttggtga 600  
 gttgcaggtc cacattcggt ctgagggtg cctgcgggtcc ttagaggcct cataggcaag 660  
 gaactcatag tttcttgcca tgaagacttt tccaac atg gct gct tac ttc ttc 714

Met Ala Ala Tyr Phe Phe

1

5

aag cct gta agg aga tcc rgt ctc agt tgg cca ggt ctt gtg aaa g 760  
 Lys Pro Val Arg Arg Ser Xaa Leu Ser Trp Pro Gly Leu Val Lys  
 10 15 20

<210> 18

<211> 23

<212> PRT

<213> Homo sapiens

<220>

<221> VARIANT

<222> 13

<223> Xaa=Gly or Ser

<400> 18

Met Ala Ala Tyr Phe Phe Lys Pro Val Arg Arg Ser Xaa Leu Ser Trp  
 1 5 10 15  
 Pro Gly Leu Val Lys Gly Lys  
 20

<210> 19

<211> 71

<212> PRT

<213> Homo sapiens

<220>

<221> VARIANT

<222> 13

<223> Xaa=Gly or Ser

<400> 19

Met Ala Ala Tyr Phe Phe Lys Pro Val Arg Arg Ser Xaa Leu Ser Trp  
 1 5 10 15  
 Pro Gly Leu Val Lys Gly Asn Leu Ile Thr Gly Val Lys Cys His His  
 20 25 30  
 Leu Ser His Val Pro Leu Val Arg Ser Pro Ser Ser Cys Pro His Ser  
 35 40 45  
 Arg Glu His Ser Pro Trp Glu Leu Cys Ala Ala Met Gln Met Ser His  
 50 55 60  
 Cys Thr Glu Ala Ala Met Leu  
 65 70

<210> 20

<211> 38

<212> PRT

<213> Homo sapiens

<220>

<221> VARIANT

<222> 13

<223> Xaa=Gly or Ser

<400> 20

Met Ala Ala Tyr Phe Phe Lys Pro Val Arg Arg Ser Xaa Leu Ser Trp  
 1 5 10 15

WO 01/40493

PCT/IB00/01507

Pro Gly Leu Val Lys Gly Thr Phe Thr Leu Gly Ala Leu Cys Cys His  
 20 25 30  
 Ala Asn Val Thr Leu His  
 35

<210> 21  
 <211> 54  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> VARIANT  
 <222> 13  
 <223> Xaa=Gly or Ser

<400> 21  
 Met Ala Ala Tyr Phe Phe Lys Pro Val Arg Arg Ser Xaa Leu Ser Trp  
 1 5 10 15  
 Pro Gly Leu Val Lys Gly Asn Leu Ile Thr Gly Val Lys Cys His His  
 20 25 30  
 Leu Ser His Val Pro Leu Val Arg Ser Pro Ser Ser Cys Pro His Ser  
 35 40 45  
 Arg Gly Phe Val His Leu  
 50

<210> 22  
 <211> 28  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> VARIANT  
 <222> 13  
 <223> Xaa=Gly or Ser

<400> 22  
 Met Ala Ala Tyr Phe Phe Lys Pro Val Arg Arg Ser Xaa Leu Ser Trp  
 1 5 10 15  
 Pro Gly Leu Val Lys Gly Phe Cys Ala Ser Leu Arg  
 20 25

<210> 23  
 <211> 71  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> VARIANT  
 <222> 13  
 <223> Xaa=Gly or Ser

<400> 23  
 Met Ala Ala Tyr Phe Phe Lys Pro Val Arg Arg Ser Xaa Leu Ser Trp  
 1 5 10 15  
 Pro Gly Leu Val Lys Gly Asn Leu Ile Thr Gly Val Lys Cys His His  
 20 25 30  
 Leu Ser His Val Pro Leu Val Arg Ser Pro Ser Ser Cys Pro His Ser  
 35 40 45  
 Arg Glu Asn Asn Ala Phe Ile Lys Ala Ser Ser Asp Val Thr Ser Asn  
 50 55 60  
 Gln Gln Ile Ser Trp Thr His

65

70

<210> 24  
 <211> 427  
 <212> DNA  
 <213> Homo sapiens  
 <220>  
 <221> allele  
 <222> 231  
 <223> 8-102-231 : polymorphic base C or T

<220>  
 <221> misc\_binding  
 <222> 219..243  
 <223> 8-102-231.probe

<220>  
 <221> primer\_bind  
 <222> 212..230  
 <223> 8-102-231.mis

<220>  
 <221> primer\_bind  
 <222> 232..250  
 <223> 8-102-231.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..20  
 <223> 8-102.pu

<220>  
 <221> primer\_bind  
 <222> 408..427  
 <223> 8-102.rp complement,

<400> 24	
cccacagaaa tatctaattg aaatattaaa ttgatttggg ttatctactc tcttgagaga	60
gatggtataa aaattcatta attataattt tatgtgcatg taacaaacag tcttttttgt	120
tttcatgggt taggtacatg tcaactgtggc aggtgtaagt gtgataattc agatggaagt	180
ggacttgtgt atggtaaatt ttgtgagtgt gacgatagag aatgcataga ygatgaaaca	240
gaagaaatat gtggaggat gtatattggc tctgtagaaa ttaataaaag tacctgtttt	300
tttcacctat ttcattgttc ttttatctct acataagttt ctgctctagg tgtgctaata	360
tagtttaaac ctatgtttat ctggaaaaca tttcacataa ggagagtggg actatggcat	420
taagagt	427

<210> 25  
 <211> 310  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 105  
 <223> 8-104-105 : polymorphic base A or G

<220>  
 <221> misc\_binding  
 <222> 93..117  
 <223> 8-104-105.probe

<220>  
 <221> primer\_bind  
 <222> 86..104  
 <223> 8-104-105.mis

<220>  
 <221> primer\_bind  
 <222> 106..124  
 <223> 8-104-105.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..21  
 <223> 8-104.pu

<220>  
 <221> primer\_bind  
 <222> 295..310  
 <223> 8-104.rp complement

<400> 25	
gattgtgttt ttccttttta gaatatTTTT gcttgaatat atgtgttttt ttcaaacaca	60
aagtaaacag aaaataataa gcgatatgtg aaatgctttc tttcrtgggt tttcttgact	120
ggattctttat aaaattcttc ttccaccta gggacttgtg tatgtgggtga atgtacctgt	180
cacgatgttg atccgactgg ggactgggga gatattcatg gggacacctg tgaatgtgat	240
gagagggact gtagagctgt ctatgaccga tattctgatg acttctgttc aggwaagggc	300
tcttccaatt	310

<210> 26  
 <211> 418  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 251  
 <223> 8-94-252 : polymorphic base A or G

<220>  
 <221> misc\_binding  
 <222> 239..263  
 <223> 8-94-252.probe

<220>  
 <221> primer\_bind  
 <222> 232..250  
 <223> 8-94-252.mis

<220>  
 <221> primer\_bind  
 <222> 252..270  
 <223> 8-94-252.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..18  
 <223> 8-94.pu

<220>  
 <221> primer\_bind



&lt;222&gt; 400..418

&lt;223&gt; 8-94.rp complement

&lt;400&gt; 26

gtggtctatg	caggtatgga	tggacattct	cacacctaac	caagcaatgt	ttgcttttcc	60
tcctcctttc	ttcccatcac	tagttgccat	gtaccgagaa	ccatctttgc	atgatgttgg	120
ggaaacggtc	ccgaagcctg	gggtgacgcc	aagtaaaagc	acaagtgcgt	ctgcaataat	180
gaatggaggc	aaaccagtca	acaagagtaa	gacaacatag	ccagatcctc	acaggtgttg	240
tgacttattc	rtcctgagca	cagttgagtg	atttatcctc	accagacatt	cctgctccgt	300
ggctgaagag	cagcaggaag	taagctaattg	cttattcttt	gctgtctccg	aacttctctg	360
ttgcaagtgg	ataaatctca	acotttgcac	ccccacaaac	aagaagacac	ctggataa	418

&lt;210&gt; 27

&lt;211&gt; 419

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 43

&lt;223&gt; 8-95-43 : polymorphic base A or G

&lt;220&gt;

&lt;221&gt; misc binding

&lt;222&gt; 31..55

&lt;223&gt; 8-95-43.probe

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 24..42

&lt;223&gt; 8-95-43.mis

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 44..62

&lt;223&gt; 8-95-43.mis complement

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 1..18

&lt;223&gt; 8-95.pu

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 400..419

&lt;223&gt; 8-95.rp complement

&lt;400&gt; 27

gactggaact	attctaggag	atttgggttt	aatagtgtgc	tcrcctcctt	gaaagcctta	60
gtgattccgg	gactctcctc	ataagtgtgc	actttcttta	tactttccag	gtaaatgtga	120
atgtggcaaa	tgcacctgct	atcctccagg	agatcgccgg	gtgtatggca	agacttgtga	180
gtgtgatgat	cgccgctgtg	aagacctcga	tgggtgtggtc	tgtggaggta	gtaacctttc	240
tcatagctgt	atgctacctg	catgcaactt	gctttacctg	cctttgttat	tctactgact	300
cttgcttttt	caattgctgt	taaaaataag	ataaacagaa	agcaatatac	cattgaacct	360
aaaagcaata	tacaattagg	tttgtaatga	cttttatttc	cctgactcct	ttatttagc	419

&lt;210&gt; 28

&lt;211&gt; 350

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

<220>  
 <221> allele  
 <222> 181  
 <223> 8-97-181 : polymorphic base A or G

<220>  
 <221> misc\_binding  
 <222> 169..193  
 <223> 8-97-181.probe

<220>  
 <221> primer\_bind  
 <222> 162..180  
 <223> 8-97-181.mis

<220>  
 <221> primer\_bind  
 <222> 182..200  
 <223> 8-97-181.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..18  
 <223> 8-97.pu

<220>  
 <221> primer\_bind  
 <222> 331..350  
 <223> 8-97.rp complement

<220>  
 <221> allele  
 <222> 98  
 <223> 8-97-98 : polymorphic base C or T

<220>  
 <221> misc\_binding  
 <222> 86..110  
 <223> 8-97-98.probe

<220>  
 <221> primer\_bind  
 <222> 79..97  
 <223> 8-97-98.mis

<220>  
 <221> primer\_bind  
 <222> 99..117  
 <223> 8-97-98.mis complement

<400> 28  
 ctgagaaacc agccattcaa tacactttta ctttggatg atgggttatt agaagcctgt 60  
 tgtcagtggtg tcaacgatat tctattcgag atggaaayct ttagttataa ggtgatggtg 120  
 agtaatagcc tttgtaattt caggttcttg tcattgtggg aagtgcattt gttctgctga 180  
 rgagtgggtat atttctgggg agttctgtga ctgtgatgac agagactgcy acaaacaatga 240  
 tggctctcatt tgtacaggtg cagtattaac cttttctaag tgctctatgc cacagtttgt 300  
 tatagagcca agagacactc gtcattgcaat gcttttagtgg gctggaccaa 350

<210> 29  
 <211> 434  
 <212> DNA

<213> Homo sapiens

<220>

<221> allele

<222> 68

<223> 8-98-68 : polymorphic base A or G

<220>

<221> misc\_binding

<222> 56..80

<223> 8-98-68.probe

<220>

<221> primer\_bind

<222> 49..67

<223> 8-98-68.mis

<220>

<221> primer\_bind

<222> 69..87

<223> 8-98-68.mis complement

<220>

<221> primer\_bind

<222> 1..19

<223> 8-98.pu

<220>

<221> primer\_bind

<222> 415..434

<223> 8-98.rp complement

<220>

<221> misc\_feature

<222> 414,416

<223> n=a, g, c or t

<400> 29

atttcattgt	ggacacttgt	gatttataat	agcatgttta	aatttggcac	attttgatca	60
taatcatrat	acctatatgt	attttattgc	agggaaatgga	atatgtagct	gtggaaactg	120
tgaatgctgg	gatggatgga	atggaaatgc	atgtgaaatc	tggcttggct	cagaatatcc	180
ttacaatta	catgagagag	gtctggattc	ttattttttc	tgggccatta	gaacatataa	240
atgcgaagga	aaccatgtat	attcaccact	aggacagggt	aaaaagacca	ttgtatgttt	300
ttctatttct	gaattacgaa	tgaatccga	gtacctatta	gaaatgagtt	atgcaaattt	360
agatgcaaat	aacattagaa	aaaaaagatt	cttcataat	taacataagt	ggtncntaac	420
gagagcaatt	tttc					434

<210> 30

<211> 477

<212> DNA

<213> Homo sapiens

<220>

<221> allele

<222> 108

<223> 99-14021-108 : polymorphic base A or G

<220>

<221> misc\_binding

<222> 96..120

<223> 99-14021-108.probe

<220>  
 <221> primer\_bind  
 <222> 89..107  
 <223> 99-14021-108.mis

<220>  
 <221> primer\_bind  
 <222> 109..127  
 <223> 99-14021-108.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..18  
 <223> 99-14021.pu

<220>  
 <221> primer\_bind  
 <222> 465..477  
 <223> 99-14021.rp complement

<400> 30  
 ttgtgtgta atcgcccctt ttctgcaaca cttgtgggtt agggaaaata attctaaagc 60  
 aagagcaaag acagagttgg gagatcacca gtgaggttca atttccrtc acattcactc 120  
 tgctccacac ctcagataat catgtgctta actgcgaaac ttgcttgaca attacagaac 180  
 actttctcac ccattactac cttgatcctc acaattctgt ggggtagtag gagcagatgc 240  
 tgaaattgcc atacgcaa at cagtgaactg aagcttagag acctccagca ggggcagagg 300  
 gtcagcgga actatcccag gggtcagcca acaagaaagt atattggaat cagagtatta 360  
 aaataagaat aataaaacca actaaaattt accgtgcttt ttatttcac tcagtgcac 420  
 caattcttaa cagtgtcagt gatggatccc tgtgcccag ggrmcagact tcttact 477

<210> 31  
 <211> 473  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 415  
 <223> 99-14364-415 : polymorphic base C or T

<220>  
 <221> misc\_binding  
 <222> 403..427  
 <223> 99-14364-415.probe

<220>  
 <221> primer\_bind  
 <222> 396..414  
 <223> 99-14364-415.mis

<220>  
 <221> primer\_bind  
 <222> 416..434  
 <223> 99-14364-415.mis complement

<220>  
 <221> primer\_bind  
 <222> 64..82  
 <223> 99-14364.pu

<220>  
 <221> primer\_bind  
 <222> 434..454  
 <223> 99-14364.rp complement

<400> 31  
 catgcccac ctaaataattt acatgccttg acagataagc ccttagcctg cccaatgggt 60  
 aacttcaaca gcatctctag tacaactggt gctcattctg tgttttaatt caaccagct 120  
 ataagatacg aatgataga attgctctag attctctatt ggtaaataa ggagatattt 180  
 gtgctattgc caataatata tgctgtacct ggataaacc ctttgggcaa gttgtgatgc 240  
 aaatactcaa gaaaataggc cacatagtta caacaggact tacctaattc cccatgggtca 300  
 tttggctgat tcagtcagtt gctttcaagc ctagggtctt ggctcaatat tattacataa 360  
 actagaattt tctattact attaatTTTA ctttgattt ttctttataa acttygtact 420  
 tattgcttgt caaatttcag cagaagtaca actcctgaga gaataatgct ggc 473

<210> 32  
 <211> 502  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 99  
 <223> 99-15056-99 : polymorphic base C or T

<220>  
 <221> misc\_binding  
 <222> 87..111  
 <223> 99-15056-99.probe

<220>  
 <221> primer\_bind  
 <222> 80..98  
 <223> 99-15056-99.mis

<220>  
 <221> primer\_bind  
 <222> 100..118  
 <223> 99-15056-99.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..18  
 <223> 99-15056.pu

<220>  
 <221> primer\_bind  
 <222> 482..502  
 <223> 99-15056.rp complement

<400> 32  
 caggaaactc acaagaagsc agatttcctt cgagcacctc ctgaataaag aggcaaaggc 60  
 cttcttaact cttacaattt acaagtggct atgagtgcyt ttatagttcc cataataatt 120  
 tctccacgta gacttcttaa ataataattt ctcctgtttt atattctctg tgcttatggt 180  
 tatatcaaac aagttaccac ttaatcaaat gccgatttgc attgctcact atgtaacttt 240  
 aattttcttt gcctcttatt tttggatctt aattctaaaa ctatgatgatc ataaattcat 300  
 ttaggaataa gcttgatgatc tagccttctt ttgaaccctt ttgtgctcct cacaatattt 360  
 gtttcgatga aacagtggac aacatttgat ctatgattgt taatagaaaa acaccaatgt 420  
 ctcaagttat tgtaaacata ggcataattg acctttgggt ctataaatat gtttggtggt 480  
 ccccaaaaata cgtctcctt tt 502

<210> 33  
 <211> 494  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 412  
 <223> 99-15229-412 : polymorphic base A or G

<220>  
 <221> misc\_binding  
 <222> 400..424  
 <223> 99-15229-412.probe

<220>  
 <221> primer\_bind  
 <222> 393..411  
 <223> 99-15229-412.mis

<220>  
 <221> primer\_bind  
 <222> 413..431  
 <223> 99-15229-412.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..20  
 <223> 99-15229.pu

<220>  
 <221> primer\_bind  
 <222> 476..494  
 <223> 99-15229.rp complement

<400> 33  
 ctgtcattga gaaatgctac caataatact tagagaattt gatacaactc agtctgaaaa 60  
 agctaagatt agcagaacag agctgtctcc aaatatattt tattaaggg 120  
 attggaccca tttttgtatg tagttccaga ggagcagatg gtgaccactg tccaggcaga 180  
 tgtgtctcaa tgtaaggaca acatctgtaa tattaataat tagaatgtat cctgtaattt 240  
 tctctctacc cttggaaacc agtcgagatc cagagtcttt cactgggagg cttaaagcct 300  
 agagcagcct tgggtctaga ggcggacagg gataatgaac taatcttgaa ccaattcatc 360  
 catagcaatc tcaatgcttt cgttagctct tataggattt taatacggcc araggatga 420  
 aggtagtctt gctggtttag aagccctgcc taccacaacc cctacaccac cccatcccct 480  
 gcatagtctg atgt 494

<210> 34  
 <211> 485  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 291  
 <223> 99-15232-291 : polymorphic base G or T

<220>  
 <221> misc\_binding  
 <222> 279..303  
 <223> 99-15232-291.probe

<220>.  
 <221> primer\_bind  
 <222> 272..290  
 <223> 99-15232-291.mis

<220>  
 <221> primer\_bind  
 <222> 292..310  
 <223> 99-15232-291.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..18  
 <223> 99-15232.pu

<220>  
 <221> primer\_bind  
 <222> 467..485  
 <223> 99-15232.rp complement

<400> 34	
caatagaaca ggctgctcct ttataattat taatcatagt gtatattaat tcatcatcac	60
atacgtggct agaaaaaat ttagaacaaa aagatatgtg atatgtaaag gacctacgata	120
attcagactt ctttgaggag agcttttatt ttattgttat tcttatttta tctcttgtea	180
atataaattg agagaataaa cagacaaaaca ttacaaatta gtgattaatt gcatttaaag	240
cctagttaaag actatttaag actattatgc ataatacagg aaaactacct ktattattta	300
tagtgggtgc cttctgaagg atctgaagga gaatcagttc tatgcctctc tcctcattcc	360
caggaggtgc ctggcattcc ttggcttgta gacgcacac cctaattctc acctctgcct	420
tcacatggtg tccctgtgtg gtgtgttttt gccccatgtg tctcctcttt ttatatggat	480
gccag	485

<210> 35  
 <211> 477  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 382  
 <223> 99-15253-382 : polymorphic base C or T

<220>  
 <221> misc\_binding  
 <222> 370..394  
 <223> 99-15253-382.probe

<220>  
 <221> primer\_bind  
 <222> 363..381  
 <223> 99-15253-382.mis

<220>  
 <221> primer\_bind  
 <222> 383..401  
 <223> 99-15253-382.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..19  
 <223> 99-15253.pu

<220>  
 <221> primer\_bind  
 <222> 459..477  
 <223> 99-15253.rp complement

<400> 35  
 aaaatcaatt ccccaacact cattttgtac gctaattttg taagatcctg aaaagtttca 60  
 ctatttttat gtttcatgtg ttacagatga aaaaaaaact agaattcaaa ttttctgagt 120  
 ttttttttac aatattttat gattacaaag ttagaagact aagaataaaa tggcctaatt 180  
 tccataatgt gagtggtaaa tgcagagcac tggcctaag aaaatatttc aaaaattag 240  
 tcactttttc cttaattttt ttccaaccta tgatctgttg aatgagcatt ttgcatatat 300  
 aaataaataa attacttttg aaataatctt gactggtttc tgttgaccac agtaaccac 360  
 tgcacagcac agcctgtaat tyctatgaac ctagggaat gtatttaagt ttattttttg 420  
 attacacagg tcctcattgt gtaactaac attgcataga atatgccagt gatgatg 477

<210> 36  
 <211> 456  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 392  
 <223> 99-15256-392 : polymorphic base C or T

<220>  
 <221> misc\_binding  
 <222> 380..404  
 <223> 99-15256-392.probe

<220>  
 <221> primer\_bind  
 <222> 373..391  
 <223> 99-15256-392.mis

<220>  
 <221> primer\_bind  
 <222> 393..411  
 <223> 99-15256-392.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..18  
 <223> 99-15256.pu

<220>  
 <221> primer\_bind  
 <222> 439..456  
 <223> 99-15256.rp complement

<400> 36  
 cctctctatg atgcttccta ttaagcaatt ggggaaatgt aataaacaag ggttggtgag 60  
 catcttcctt agtgagatgt ttttgaaga attggataat tgagtgaata atagtgaaga 120  
 actcctgtgt ctgatgttgc tccatgttgg aatgctttta tgttctcaga gaatgagtc 180  
 ctgagagcca attgtgatga tacacaatgg ttttaccag gttggatatg gtcctctgta 240  
 ctggtaccct ttaagtcagt ggcactaatc agtcagtcac tgcacgctt tgtgttggtc 300  
 catcatatgg tatgccctct tagagaacat cctgattagt ccttagacat cttttcaatt 360  
 tgaacactgg ggctcctcat tcgggtaaaa aytatggaca gtcagtgaag ctgttgcaat 420  
 ggccctcat agcagattgg atctcaatgc actttg 456

<210> 37



<211> 541  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 337  
 <223> 99-15258-337 : polymorphic base G or T

<220>  
 <221> misc\_binding  
 <222> 325..349  
 <223> 99-15258-337.probe

<220>  
 <221> primer\_bind  
 <222> 318..336  
 <223> 99-15258-337.mis

<220>  
 <221> primer\_bind  
 <222> 338..356  
 <223> 99-15258-337.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..19  
 <223> 99-15258.pu

<220>  
 <221> primer\_bind  
 <222> 521..541  
 <223> 99-15258.rp complement

<400> 37  
 cagtaaagaa ggacttagga aatatttttt gaatcatgag ctgctaatag atcagtaaat 60  
 aggatttatt aaacacagct gtattcagga ctctgacctg tgtgggcctt agggaaatac 120  
 aaaatttaca aactgatgtg attttcttac cagtaagaaa tcttatgaaa gattaaatta 180  
 aaaaaatata aatgggggca actcaatcac ataaaagtaa tatattccat gtgactgaaa 240  
 tactagggaa aatttagaag tagaatttaa gtgagaattt caagcagggt tctgaaagat 300  
 tctacaggat gaaggaagaa ttgcccatat gccaatkgcc tatgtgaaat cacaaaaaca 360  
 gaaatcaata aggcattgat ggaagtacag agtgattgcc tagttagaaa aagagcgtga 420  
 agggatcagt taacacaaac aagaagagtt taatgcttat ctgcaataag ttggtagagc 480  
 aaaattaagc acttcccagt atactgatag acgttaatty coactctacc atgtgaaatt 540  
 t 541

<210> 38  
 <211> 500  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 200  
 <223> 99-15261-202 : polymorphic base A or G

<220>  
 <221> misc\_binding  
 <222> 188..212  
 <223> 99-15261-202.probe

<220>  
 <221> primer\_bind  
 <222> 181..199  
 <223> 99-15261-202.mis

<220>  
 <221> primer\_bind  
 <222> 201..219  
 <223> 99-15261-202.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..19  
 <223> 99-15261.pu

<220>  
 <221> primer\_bind  
 <222> 481..501  
 <223> 99-15261.rp complement

<400> 38  
 cttctaattcc ttgtttcca cttattttat ttcattcctc attttatccc ttttttctaa 60  
 attccattttt attatactta aggtgctttt aatatgggta tcatactcct gatagtgtta 120  
 tttctttctt agtcttctta tataagcgct atacgttcac attccatctc ctttggttat 180  
 ctttcatttt cttcaccgar cctctttgct ctcttttttt atagctgggt cactcaaaat 240  
 gtcttacttt gccatttttt aaatttattt tcattctttt atgtactgaa taaaatttaa 300  
 aaatacttta tcatgggtggg aggtaccggt gatgtccaaa taagtgttta tattaattgt 360  
 tgggggttttt ttgtttgtgt gttttttgaa aggttaagaa aatctcattc agaaagtaag 420  
 ttgttttaaaa attctggacc aaatttacca cacatcaagc agatacttac caagttgttt 480  
 gtagacatta gcagtattta 500

<210> 39  
 <211> 541  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 432  
 <223> 99-15280-432 : polymorphic base C or T

<220>  
 <221> misc\_binding  
 <222> 420..444  
 <223> 99-15280-432.probe

<220>  
 <221> primer\_bind  
 <222> 413..431  
 <223> 99-15280-432.mis

<220>  
 <221> primer\_bind  
 <222> 433..451  
 <223> 99-15280-432.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..18  
 <223> 99-15280.pu

<220>  
 <221> primer\_bind  
 <222> 521..541  
 <223> 99-15280.rp complement

<400> 39  
 atgtccatcc atcttgccca gagagagttt ctacaacact tctctgcaa gccctttccc 60  
 tacttgctc acctattgct ttcctctgtt acgttgatt cccctcactg tttcttccaa 120  
 catcttccca cctcagagca tggacacttg ctgctctttc tgtgtcatga tgctgtcac 180  
 ttgtcccttt cttaatgtct cctccctgag ccaatcttct ccacccccac aacttacgca 240  
 cacttacatg tcatattttc cttcatagcc tttaacacca tttgaaatga tatatatattg 300  
 attgctttta aaattttctc gtccccccac taaatataaa cttcaggatg gcaagaatgt 360  
 agtccattat cttattttctc cagcctccat acttttaaga aaataaattt tggttgtata 420  
 agccatccag tyagtgtac ttggttatag caccctagc aaaagaatac aaaaaaaggg 480  
 agaatgtttg caatcatctg ttgaggcta ggaattccca gagagggraam caaggagtaa 540  
 t 541

<210> 40  
 <211> 489  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 150  
 <223> 99-15355-150 : polymorphic base C or T

<220>  
 <221> misc\_binding  
 <222> 138..162  
 <223> 99-15355-150.probe

<220>  
 <221> primer\_bind  
 <222> 131..149  
 <223> 99-15355-150.mis

<220>  
 <221> primer\_bind  
 <222> 151..169  
 <223> 99-15355-150.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..18  
 <223> 99-15355.pu

<220>  
 <221> primer\_bind  
 <222> 471..489  
 <223> 99-15355.rp complement

<400> 40  
 taactttctcc gtctctcctt cttagcccat atgtcaataa tgactgaaag tattcatttc 60  
 catcttttaa ctgcctatc cagccacctc ccacctccat ctctttcctt ctaagttttc 120  
 ttcattctct actttgggca aaaggaaaty gatgtgtcag acaggcctag ttttgaattc 180  
 tggatctgct agcacttctc tgtgtgtcct tggttatatg atatagtctt aaaccttaat 240  
 gttcttgctt gtaaaatggg gataataaaa acctctaac agtggttggt tcatgcagct 300  
 ttcattacaa acttctcat tcaaaatctt caatgatctt catttttcac aaaatgaaat 360  
 tcaaaatttc ttagattat tgagacaagt cccctactct tcacctaaat ttatctttta 420  
 tttattctct catcattatc aacaactact aggccttggt gccttgactc cagaggcaaa 480

aatctttatc

489

<210> 41  
 <211> 450  
 <212> DNA  
 <213> Homo sapiens  
  
 <220>  
 <221> allele  
 <222> 298  
 <223> 99-15663-298 : polymorphic base C or T

<220>  
 <221> misc\_binding  
 <222> 286..310  
 <223> 99-15663-298.probe

<220>  
 <221> primer\_bind  
 <222> 279..297  
 <223> 99-15663-298.mis

<220>  
 <221> primer\_bind  
 <222> 299..317  
 <223> 99-15663-298.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..18  
 <223> 99-15663.pu

<220>  
 <221> primer\_bind  
 <222> 430..450  
 <223> 99-15663.rp complement

<400> 41  
 tccacacctt ttctaaacgt gttgcttcaa tacgttgata ggtgaggaca cttaaaaatt 60  
 agactttata gaaatagggt tttttttgtt tacatatata gttcttttgg tatcatatat 120  
 ttagcctctt tctaaaattt attttttgat actgaaggga gaaataggga gttattaatc 180  
 aacaggcatt aatttttagtc aagcaaaata aataagctgt agcgatctgc tctgtaacat 240  
 tgtacctaca gccacaatt atattgtgtc cacttaaaaa tgtgttagat ctcatagyaa 300  
 ctcttcttac cacaataaag taaaaattct gaaacaataa gtgaatacct aaataatata 360  
 aacaaatata atattgtagt tttgggcact taataaatga cagcctcatt tctcaattag 420  
 agatcatcac aagttagaca gatgacgatg 450

<210> 42  
 <211> 502  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 185  
 <223> 99-15664-185 : polymorphic base G or T

<220>  
 <221> misc\_binding  
 <222> 173..197  
 <223> 99-15664-185.probe

<220>  
 <221> primer\_bind  
 <222> 166..184  
 <223> 99-15664-185.mis

<220>  
 <221> primer\_bind  
 <222> 186..204  
 <223> 99-15664-185.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..19  
 <223> 99-15664.pu

<220>  
 <221> primer\_bind  
 <222> 483..502  
 <223> 99-15664.rp complement

<220>  
 <221> misc\_feature  
 <222> 54  
 <223> n=a, g, c or t

<400> 42  
 gtttaccatt agcactgtca tatttgtgtg acttgtcatt ctctacagcg gagnacgggc 60  
 tggcacgggg cctgatgctg acttgcacaa gggaagcctc ctgtctctga cttccccagg 120  
 ataattcctg gggaaagtgt gctccctagt gttaagagcg gtttaatggc tggagggttt 180  
 cagckggctg accaggcaga gaaggagggt gaatcacctc tcagcactct ccacttagac 240  
 tttgtgtggt cgtcgggttg tcaaaccttc taactagtgt tattgcagat ttggcattcc 300  
 agtgcaaaaca aaagacagaa acacaatgtt cacatgcttt ccagagatca cctggatata 360  
 agatcatttg attttcaagt aagtcgaaac cttgggtggaa atcattaact atcctgttta 420  
 tgaccacaaa ataaaatccc aaatttcctc tcttcatttc ttacctgctt taaaattgta 480  
 tccaaagcgt gaaagtaaaa ga 502

<210> 43  
 <211> 476  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 398  
 <223> 99-15665-398 : polymorphic base A or G

<220>  
 <221> misc\_binding  
 <222> 386..410  
 <223> 99-15665-398.probe

<220>  
 <221> primer\_bind  
 <222> 379..397  
 <223> 99-15665-398.mis

<220>  
 <221> primer\_bind  
 <222> 399..417  
 <223> 99-15665-398.mis complement

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 1..20

&lt;223&gt; 99-15665.pu

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 458..476

&lt;223&gt; 99-15665.rp complement

&lt;400&gt; 43

cgtaaagtgtg aaaagcatag cctcttcttg gaatgttaag tataaatatc tgaaatactg	60
ggcttgatat gtcaacagga gattgatgga taaaaataga attttatata aaaaacaact	120
ggacatatta gattgttaac ttggaagaaa gaccatattc aaagaagaaa acatagtgc	180
taatttcaaa catttaaagt ctccctgtg gaaacaaagg aatatcttg ttctaact	240
tcaaagaaca gggttaaaaa atagactcac cacagagtaa atgcacaatt gacaatcgtg	300
aatgaattaa aaaccaaaca aaatatcttg tcagctttct atctatgaaa ctaagaaaca	360
ggcttctac taaggtaatg aatgtaattc acagagarca ttcacgtata agtttcattc	420
atgtttcaaa tttcattgat ttgatcaatg gggtattcta ataccctccc ttattt	476

&lt;210&gt; 44

&lt;211&gt; 459

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 139

&lt;223&gt; 99-15668-139 : polymorphic base C or T

&lt;220&gt;

&lt;221&gt; misc\_binding

&lt;222&gt; 127..151

&lt;223&gt; 99-15668-139.probe

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 120..138

&lt;223&gt; 99-15668-139.mis

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 140..158

&lt;223&gt; 99-15668-139.mis complement

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 1..18

&lt;223&gt; 99-15668.pu

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 439..459

&lt;223&gt; 99-15668.rp complement

&lt;400&gt; 44

agaggtgaga gaaaggtgaa ttgtgcaa atggacatcg aagttattca gacacatatg	60
gtttccatgt cgtgagacct taaagcactc tcagaaactc caagaaagct gttgattggt	120
cagatttctc acttttctgc cctcaaaaca agattgtcct gaacttatag gaaatatact	180
caaaagtagt atgatttaag gaagtcttcc aaatacctgc ctggtttaca gataatcaat	240

```

ggaattccaa aagattttat gtgttttaag ccctaaaaat gcaaatatcc aggaaagggc 300
tagactaaa cagagctggc atgtatgcct cagtgcaggt gtatatctca gtgtctcctg 360
aagccagggc cttagaggat tgtggaaata ggtacactga cagcagggac tggaagccac 420
agtctggaga gcaaagagca gaatccatac ttacacctc 459

```

<210> 45  
 <211> 547  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 166  
 <223> 99-15672-166 : polymorphic base C or T

<220>  
 <221> misc\_binding  
 <222> 154..178  
 <223> 99-15672-166.probe

<220>  
 <221> primer\_bind  
 <222> 147..165  
 <223> 99-15672-166.mis

<220>  
 <221> primer\_bind  
 <222> 167..185  
 <223> 99-15672-166.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..18  
 <223> 99-15672.pu

<220>  
 <221> primer\_bind  
 <222> 533..547  
 <223> 99-15672.rp complement

```

<400> 45
ccaataccat aactcctcta taggacatgg aagagtatta tatatgacaa atgattgcta 60
tgattattat tatcagtgtt attattatcc taatcctaag taatccaata aaagaaaaat 120
acatctgtgc ctgtgcgtat gtgcacgtgt gtgcagtcac atacaygttg agtaaaggta 180
aagtctagct gtatttaatc aacctacctg aatcctcagg aaaaaattct aaacctagtt 240
taaaacatgt aaactctaag ctctctcctt atagtcagtt agtagcagca catcttaaaa 300
tctgggtgtga atattctctt agttctacat gagtctaact aaacagagga ttattcttag 360
gtgtttgaaa gagacatatg tgacactgct gttttgagaa caatttaagt gttgtcttgt 420
catgtacaga agttctcata ttactttata taaatggttg cataattggt ttatagtaaa 480
taatagactg tcaatatttc taggataact caaaacaaa atttcctaga mmacattttg 540
aaaaggg 547

```

<210> 46  
 <211> 500  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 318  
 <223> 99-15682-318 : polymorphic base A or T

<220>  
 <221> misc\_binding  
 <222> 306..330  
 <223> 99-15682-318.probe  
  
 <220>  
 <221> primer\_bind  
 <222> 299..317  
 <223> 99-15682-318.mis  
  
 <220>  
 <221> primer\_bind  
 <222> 319..337  
 <223> 99-15682-318.mis complement  
  
 <220>  
 <221> primer\_bind  
 <222> 1..19  
 <223> 99-15682.pu  
  
 <220>  
 <221> primer\_bind  
 <222> 482..500  
 <223> 99-15682.rp complement

<400> 46  
 gtttgatgag caacaagtgg ccacacaagt attggagaaa aaaaaaatta gaatgtgggg 60  
 agagaggcag tagacacaat aaaaaattct atataacaaa tgctatatct ttaagaaatc 120  
 gaatctttcc attaaatatt tgtttaaagt tacagactac ttatagtgta aagtatcact 180  
 tatcacaag atcagtaact tatttcctct gtatagtaga tgccctctct gtagaattaa 240  
 catttcacta taatctctaa gatacagttc aagggcacaa ttttataatt caaagatatt 300  
 tacggcatga aattgagwtc taaaactott gtatctttta caaccathtt gtaattaaat 360  
 aactgttgac atttacaatc attgaagaac tatttgaact ctgtagagta tagatattga 420  
 agtacaaaat taattatgag ataatatatg gagcctgkcg gctttatttt ttaataagtw 480  
 tgtatgtatg ccaataaagg 500

<210> 47  
 <211> 461  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 217  
 <223> 99-16081-217 : polymorphic base C or T

<220>  
 <221> misc\_binding  
 <222> 205..229  
 <223> 99-16081-217.probe

<220>  
 <221> primer\_bind  
 <222> 198..216  
 <223> 99-16081-217.mis

<220>  
 <221> primer\_bind  
 <222> 218..236  
 <223> 99-16081-217.mis complement



<220>  
 <221> primer\_bind  
 <222> 1..18  
 <223> 99-16081.pu

<220>  
 <221> primer\_bind  
 <222> 443..462  
 <223> 99-16081.rp complement

<400> 47  
 caaaggacag agaaccagca ggcaggccct gggtaagcc ccgtgatgtc ccaggaggag 60  
 gggaggagg cggagtcgga agggagcgag ggagccagg aggcagggcc ctctgagggc 120  
 agggagttct ccatgaagtt ccagaagcca atgaccagat cggcgcatgc aaacagctta 180  
 aaagaaaaca ttaggcactg gtggcttcgc tctaaaygaa cagttcttaa aactgctttc 240  
 agttctagcc ataataaagg catTTgaatt atgttgcttt tgatgaaact ataaagatat 300  
 tggtttttat acagtatcat gccagaaaca aatttataac attgttatgg tataacagca 360  
 ggaatacttg gaatacacat ttttcagctg agctctaaga attacttttc ccttgcccca 420  
 ggttggaat taagaactat aacttactcc acccactcc a 461

<210> 48  
 <211> 532  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 218  
 <223> 99-16082-218 : polymorphic base A or G

<220>  
 <221> misc\_binding  
 <222> 206..230  
 <223> 99-16082-218.probe

<220>  
 <221> primer\_bind  
 <222> 199..217  
 <223> 99-16082-218.mis

<220>  
 <221> primer\_bind  
 <222> 219..237  
 <223> 99-16082-218.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..18  
 <223> 99-16082.pu

<220>  
 <221> primer\_bind  
 <222> 512..532  
 <223> 99-16082.rp complement

<400> 48  
 caattcagca agagccactt tggttaacaa accttotggg aatgaagggtg ctttacaaag 60  
 aagcaggctg tgataaaaac aaacaagaca acattcataa actgtgggaa gcattgttgt 120  
 cattgggtca cactcctgc accaggcagg tggttgtctg ctgcgttcct ttacagaaca 180  
 gatgctgagc ctgcctggag gtggcaattt accaaterat ggtcttaagc aattccaaag 240

```

gctggggtaa ggggtagaaa agagaagctc aaaaatgtta gggatggaga aggggggagt 300
gattcaggaa gaggagaaac ctgttcattt cctcccaaact tacaaccact gtcacatcca 360
tggttagtcc tgtgcaaaact tccagtcctc agctgatatc cctgcagaca agtagagaga 420
ggctagaaca aaagctaaag tgtagatgtc ccatacacta atcagtgttg ttttttaac 480
agaggttgaa attcatactc tgatgcaaac tgtattagtc agwtgttgct gc 532

```

<210> 49  
 <211> 451  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 81  
 <223> 99-20933-81 : polymorphic base A or C

<220>  
 <221> misc\_binding  
 <222> 69..93  
 <223> 99-20933-81.probe

<220>  
 <221> primer\_bind  
 <222> 62..80  
 <223> 99-20933-81.mis

<220>  
 <221> primer\_bind  
 <222> 82..100  
 <223> 99-20933-81.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..19  
 <223> 99-20933.pu

<220>  
 <221> primer\_bind  
 <222> 432..451  
 <223> 99-20933.rp complement

```

<400> 49
ttccattcac atgagaaact atgtgtttac tatttctggt atgttttgta tgattttgtg 60
tgcttgata ttttcagtaa mtgctggggt agcaagtgtc agaatcactc caaaaaaact 120
gagatatatt tgccaagtca acagtctcca gactattggt tcagcaagac aaggggcaag 180
gctctttctg ggcctgagta atcagtaaat atataaaaag aataagaata atagtgaagt 240
tactgctaga cgtatttctt tacatctgaa cattttcttt catttgcaac acttaacctg 300
gaatcgggtg cttccagaaa atcctgtaaa aatcatgtag cttccaagat ttaaagctgg 360
cctagawaac atgaacatga tcccatgtct tatctttact catacatcag gaattctcta 420
aatttctggg agaacatttt gtagaagaca c 451

```

<210> 50  
 <211> 511  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 71  
 <223> 99-20977-72 : polymorphic base A or C

<220>  
 <221> misc\_binding  
 <222> 59..83  
 <223> 99-20977-72.probe

<220>  
 <221> primer\_bind  
 <222> 52..70  
 <223> 99-20977-72.mis

<220>  
 <221> primer\_bind  
 <222> 72..90  
 <223> 99-20977-72.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..18  
 <223> 99-20977.pu

<220>  
 <221> primer\_bind  
 <222> 491..511  
 <223> 99-20977.rp complement

<400> 50  
 gtacatggta ggtggtagat ggtgaaataa tttttatatt taaggagttt caaagatgat 60  
 ctaaaacatg maggcctaaa acaatagatg aaacactatt tcaaaatgta ttctctcttt 120  
 gttttcactc tccatgttac aaaaatggca acaattacg taaatgccaa ggaagccaat 180  
 atatcatatg gtaaatgaca aatcaactgt aacattgaaa caacaattaa tagctattat 240  
 agctattaaa atttgaagca aaattgtgca tctttaaaaa ttatacacac atgtgagagg 300  
 actcatgact gaaatctcaa tataccatct atccaagttg agttattttc cttttttttt 360  
 ttttttaaga taattattct tcttctact tctgggaggg acagtgtatg ggaggtacct 420  
 gtcaggcact gtgagaaccc tgggctctga acagatatgg ggtattttct tcagtatcct 480  
 caatgcagct ccaataacta atcccatttc t 511

<210> 51  
 <211> 466  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 89  
 <223> 99-20978-89 : polymorphic base G or C

<220>  
 <221> misc\_binding  
 <222> 77..101  
 <223> 99-20978-89.probe

<220>  
 <221> primer\_bind  
 <222> 70..88  
 <223> 99-20978-89.mis

<220>  
 <221> primer\_bind  
 <222> 90..108  
 <223> 99-20978-89.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..19  
 <223> 99-20978.pu

<220>  
 <221> primer\_bind  
 <222> 446..466  
 <223> 99-20978.rp complement

<400> 51  
 acttaactgc tacctaattgc agcctgtaca agcggataaa ttgaactaaa atgtcttcaa 60  
 cactaaaaat aaaatgcaaa tcaaatctst ttaaaaaacc aaaaccaact tgtattttga 120  
 agattaactg ccaactgtgtc taaacccaaa tctgtctagc cagagctgaa tttgtgatta 180  
 caaagcagta acgcaaagggt tagaaaggat caaagatctt gaaaaacctc ttctcagatc 240  
 catctcactt ttccattggc cacattttcca atacatctct gcccattaa aaccttacca 300  
 aaattgctag ttcaagtacta taagtttaaa acaaatgggt tgctcaattc cttttgatgc 360  
 atcttcgcga tctctcgctc tttttctttt tttctatttt cactgccagc ccttcattac 420  
 ctctcttagg tggctctccc acctgcagtt ttccttttca aatcct 466

<210> 52  
 <211> 449  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 300  
 <223> 99-20981-300 : polymorphic base A or G

<220>  
 <221> misc\_binding  
 <222> 288..312  
 <223> 99-20981-300.probe

<220>  
 <221> primer\_bind  
 <222> 281..299  
 <223> 99-20981-300.mis

<220>  
 <221> primer\_bind  
 <222> 301..319  
 <223> 99-20981-300.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..18  
 <223> 99-20981.pu

<220>  
 <221> primer\_bind  
 <222> 429..449  
 <223> 99-20981.rp complement

<400> 52  
 tttgccatcc ctttgctgat gggttgattt ggaatgtgtc atttogtttg tgaaatgaga 60  
 catagtcaag gtacccatgt gctcaagtgc ttagattaca acaatgtttt gatgcaaag 120  
 tgcaaagtac tcagagatca acactatggt ctaataacag gccatagcaa aatagagcaa 180  
 atgaatcctg tgaaatcaga gagaatgcat caaataaaga tgtactacct aaagtgcata 240  
 atctactagg agcactttga catttcagtt caagttctac agtggtttcac attttttcar 300

tctctattgt agtctcactt ggaggtgatt gcttgatcac tagaggaata tgatattatt 360  
 cgtttcagaa atgaagcagc tatggtgttc tcttgtgaat agcactagat atgaattagg 420  
 attatgatct tgtttgtgta gaatgttcc 449

<210> 53  
 <211> 449  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 48  
 <223> 99-20983-48 : polymorphic base A or G

<220>  
 <221> misc\_binding  
 <222> 36..60  
 <223> 99-20983-48.probe

<220>  
 <221> primer\_bind  
 <222> 29..47  
 <223> 99-20983-48.mis

<220>  
 <221> primer\_bind  
 <222> 49..67  
 <223> 99-20983-48.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..19  
 <223> 99-20983.pu

<220>  
 <221> primer\_bind  
 <222> 431..449  
 <223> 99-20983.rp complement

<400> 53  
 catgacacat caataagaga agcctgctgg cacagaggag ggcaggtraa accacctttg 60  
 tagaagaggc tcatttgaaa atgctcctga aagatatgat atcagaacac tatcttatgt 120  
 agagtggata agactttgat gatcttgaaa atgcttactg aggaatacat agcattttatc 180  
 tttctttcct tgattaaaga actttcatat atacatattt atcctgtatg gataggtagg 240  
 agaggtgatg ccagccocaca gagatgacca gtccctattc cagaagttct gagttgttgc 300  
 ttaagtgaag aagtatgctg gtctgcttcc aaacttctgt aaacttcagc ttagtggttag 360  
 aacagttaga aaatccaaat atttagtttc aatgcctgag tagtccattt actgattttc 420  
 accaactctg taccatcacc ttgtgtctaa 449

<210> 54  
 <211> 466  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 148  
 <223> 99-22310-148 : polymorphic base C or T

<220>  
 <221> misc\_binding

<222> 136..160  
 <223> 99-22310-148.probe

<220>  
 <221> primer\_bind  
 <222> 129..147  
 <223> 99-22310-148.mis

<220>  
 <221> primer\_bind  
 <222> 149..167  
 <223> 99-22310-148.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..19  
 <223> 99-22310.pu

<220>  
 <221> primer\_bind  
 <222> 446..466  
 <223> 99-22310.rp complement

<400> 54  
 gtgtgatacg agatcattct ttttttaaaa gctccctagg gtttcagaag tgcagcacag 60  
 gttgaaagct accgatatag agcaaagctt tactaaaagc tgatatgtgg aagcaaagaa 120  
 tgtgaaatga ctctctgaagg caccaaaytg acaatgcaaa tcaaatccaa gaacaagctc 180  
 taaatgtgtg cttgtatttt actgaatccc ttttgtaaac ccatgcaata ttttacccca 240  
 atattattgc aatacaaaact gggaaaagata agatactgga atgtaacctg catcagacag 300  
 atagcaatat ctcatgtgatt tagagcagag cagaccaagt tttggagtag agggagtagt 360  
 togatgactg gcagagaagc gggtaagaat ttaaagacaa tagtagaatc attatctctc 420  
 tgctacctcg ttgggacaaa gctctctaaa ataagcaact acacag 466

<210> 55  
 <211> 471  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 147  
 <223> 99-25922-147 : polymorphic base G or T

<220>  
 <221> misc\_binding  
 <222> 135..159  
 <223> 99-25922-147.probe

<220>  
 <221> primer\_bind  
 <222> 128..146  
 <223> 99-25922-147.mis

<220>  
 <221> primer\_bind  
 <222> 148..166  
 <223> 99-25922-147.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..21

<223> 99-25922.pu

<220>

<221> primer\_bind

<222> 451..471

<223> 99-25922.rp complement

<400> 55

ctctaataatc aataaagcct caaacatggt tgctgattct aggacatatt gtcaatcacg	60
cactcacagg agaattttcc atctacctga aataatttaa atttatctta tatttctctt	120
tgaigagaat catacaaaat ggaattkac tgaattcttg tgtttgaga gcacaacata	180
tagcagaact attaaaaaaa atcaagtaca tttgtaatat ttgaaataga aaagaagttg	240
gccaaagatat ccccaaatat gctaaaaagt cctctaattg tgatgctgaa agaaactttt	300
ctaaactatc aaaaatgaaa tgcaaatgt agttagtagg gctaaaaaaa ttactaaatt	360
acctttttct tatctctaca aaaaacatta caaataatat tttgtcgtat gaagtgtgat	420
caaagtatat gtaactaaaa actaaggaat catttatagt ttatcaggca g	471

<210> 56

<211> 499

<212> DNA

<213> Homo sapiens

<220>

<221> allele

<222> 121

<223> 99-25950-121 : polymorphic base G or C

<220>

<221> misc\_binding

<222> 109..133

<223> 99-25950-121.probe

<220>

<221> primer\_bind

<222> 102..120

<223> 99-25950-121.mis

<220>

<221> primer\_bind

<222> 122..140

<223> 99-25950-121.mis complement

<220>

<221> primer\_bind

<222> 1..19

<223> 99-25950.pu

<220>

<221> primer\_bind

<222> 479..499

<223> 99-25950.rp complement

<400> 56

ctaccaaatc atcaactacc aaatacctca gtgcattttc tagcattaac tttatttagt	60
accagaatta tcacagagtc taaagttatt tcacacttaa tttttagaaa ataaatagat	120
stgtagaaat ctctcttatg gattcaatat tttgggaaag aagcctcagt gatggatatt	180
ttgttggtt agaagtggct atctattgtt actttgtaat aaattatctt aaagggtgtt	240
tattgtaagt atcacatata atccaggata ttttcttttg gaagaggcaa acccagttaa	300
attgggaaaa agaatatagt ctgcaaatac aaacattgaa aaatgtgacc tctgcccctt	360
ttcaactctt ataggacaaa ttcagaactc aactaaatga aattctttat attgttaaaa	420
aatttcccct agatctcaga cagatgctga gatctgtttt gtagcattca cccacgtacc	480

taaattctcac tctcagcaa

499

&lt;210&gt; 57

&lt;211&gt; 484

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 373

&lt;223&gt; 99-25961-376 : polymorphic base A or C

&lt;220&gt;

&lt;221&gt; misc\_binding

&lt;222&gt; 361..385

&lt;223&gt; 99-25961-376.probe

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 354..372

&lt;223&gt; 99-25961-376.mis

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 374..392

&lt;223&gt; 99-25961-376.mis complement

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 1..20

&lt;223&gt; 99-25961.pu

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 464..484

&lt;223&gt; 99-25961.rp complement

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; 395

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 57

tctcccctgt	gaatgggaac	atgtatgaat	ttgaaattat	gttctacatt	ctcctccaca	60
gttggaaaag	gagggaaaga	cagacacacc	cctggactca	gagggttctg	cctgcctcac	120
caagggtctg	aagggtcact	ttctgecttc	tttcccttcc	attctttctt	caaataattg	180
tccatgaaac	tatatgcagt	tggagcctga	tggctcttcc	cttaccacct	cccccaagtga	240
aaatggkttc	cttggttgaat	ttgtgagaag	cctttcccca	agtctcttgc	tttgcttgaa	300
tcccagagct	tagatttccg	agaaaataaa	ctatgcagta	ggaataggat	tgggtgtggct	360
cctgggacg	ggmtacacag	acaacttata	tctcnaaact	ctggaagacg	atcattgtaa	420
acccttgaat	gtggcagggg	catggcaggt	cactgtgaga	ttagcttctg	ttaaactgag	480
ggct						484

&lt;210&gt; 58

&lt;211&gt; 470

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 399



<223> 99-25965-399 : polymorphic base A or G

<220>

<221> misc\_binding

<222> 387..411

<223> 99-25965-399.probe

<220>

<221> primer\_bind

<222> 380..398

<223> 99-25965-399.mis

<220>

<221> primer\_bind

<222> 400..418

<223> 99-25965-399.mis complement

<220>

<221> primer\_bind

<222> 1..21

<223> 99-25965.pu

<220>

<221> primer\_bind

<222> 451..470

<223> 99-25965.rp complement

<400> 58

ctcaaaatac	atatacacca	gctttaaatg	ggaggccaag	ataacggaaa	gaaagaatga	60
ggtttcgatt	gaggatgctt	agataagaaa	agcgagagc	cccctttcct	gggtcaagtt	120
gcttttcagt	gccagctcac	ctttgtcaac	ttcctaggtc	aaatgtcatg	ggaaaggaaa	180
ataatgatcc	aatcagtaaa	ctttaaatga	tagaataaat	agagaaaaca	attggctgag	240
agaaagctga	ctcgattgta	acgttttagct	actagagaag	atgaaagttc	aaagaactta	300
acatgaagtg	taaggagata	ataagttttg	ctcattgtta	atttgtctca	tgattctgtc	360
aatgcgtact	ggtcacgtga	ctttcacaag	gaaaatgcrt	tattcaccac	agcagatcat	420
cttcctcatt	aaaaaccaa	ggtacaatgt	gggtactgtg	acttttagaa		470

<210> 59

<211> 523

<212> DNA

<213> Homo sapiens

<220>

<221> allele

<222> 241

<223> 99-25966-241 : polymorphic base A or G

<220>

<221> misc\_binding

<222> 229..253

<223> 99-25966-241.probe

<220>

<221> primer\_bind

<222> 222..240

<223> 99-25966-241.mis

<220>

<221> primer\_bind

<222> 242..260

<223> 99-25966-241.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..21  
 <223> 99-25966.pu

<220>  
 <221> primer\_bind  
 <222> 503..523  
 <223> 99-25966.rp complement

<400> 59  
 caatagtgtgta aacagccaac cattaataaac acataaaata agtttcataa agatgaaggt 60  
 gacacaggag tccagactaa aawgtatgag tatgagttca ggaaagggtt tctaagaagg 120  
 tattattttaa cttctgggca ttcataattta agtataggaa aagtcatgtc agtattcagg 180  
 aaaaatctga gtcattgcttg atggtatgaa agtgtgaggc atgtattgag aatgttcaag 240  
 rgctgggttga ctgccactcc ttctctccaag ctcatagcaa atcccacttc agcacggaac 300  
 tgttctttaa gtccacactt accgctcttg ctccctcatt ccccatgtgc cttgtttatt 360  
 tagaggtaga tatattggtc aggtcaggca agctaagctt cagtaacaca tcaaccctta 420  
 catttcagtg gcttcaccca gtgaagattt gtttctcact taggctccaa ggttatctta 480  
 ggctggcgagg agatcagctg cacatcatct tcaactgactc tca 523

<210> 60  
 <211> 493  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 56  
 <223> 99-25967-57 : polymorphic base A or G

<220>  
 <221> misc\_binding  
 <222> 44..68  
 <223> 99-25967-57.probe

<220>  
 <221> primer\_bind  
 <222> 37..55  
 <223> 99-25967-57.mis

<220>  
 <221> primer\_bind  
 <222> 57..75  
 <223> 99-25967-57.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..20  
 <223> 99-25967.pu

<220>  
 <221> primer\_bind  
 <222> 473..493  
 <223> 99-25967.rp complement

<400> 60  
 aacaagtgc atgacaaatg tacatcttct ctatcttgac tggtaacacag agcctrgatt 60  
 gaaaaggagc aagagaagaa gcaggcaca gagaacagct gggctattac tataattgga 120  
 ggtgtgagaa agagaaggtc aagaactcct ccagttttc tggcaaaaac aagttaaaaa 180

cataaaaaaa	taaaaaaata	aattaaaagc	aaataaatgt	gccaacgaca	aagtcagaga	240
gtcctagaag	gacaagctcg	tgagggatag	gaagggtcaag	ggttttatatt	cattgtatta	300
agaatggaga	ttttgctgag	ccataccgca	cagatatcta	gtcagttgga	aacataaatc	360
taagtgaaga	acaaagttct	gtgctagaga	taattattca	agaaccattc	ttaaacaaaa	420
tgattatact	agaaaagata	tatcaattac	ttttcatata	aaacaaaaaa	tgcgatattt	480
ggcattctga	ttg					493

<210> 61  
 <211> 530  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 200  
 <223> 99-25969-200 : polymorphic base G or T

<220>  
 <221> misc\_binding  
 <222> 188..212  
 <223> 99-25969-200.probe

<220>  
 <221> primer\_bind  
 <222> 181..199  
 <223> 99-25969-200.mis

<220>  
 <221> primer\_bind  
 <222> 201..219  
 <223> 99-25969-200.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..21  
 <223> 99-25969.pu

<220>  
 <221> primer\_bind  
 <222> 510..530  
 <223> 99-25969.rp complement

<400> 61	
gttagtgact ccttctttta cacatttctg gtttcaaagc agattatcaa gtaaaatttt	60
aaactgccac caggagattt agctaaggag taccagctga ttttgttcta atgcagtctt	120
taaaataatg tacactgtga catattcaag tcaagtacag catttatatg ctacacagtc	180
aagtaaaactg tgtagcatak aaatcctacc acattctcca aaccagacag agggagaatc	240
acagatgccc catgaatcac ctagaaatcc tgtcaacata taacaaatta tatgtggtag	300
accataaaag tcaatatcaa gttaaaagca ttttttttgt cctcatgtaa aataaaactct	360
ggaaggagag aatgatcttc acatgtcaat gaaaatctga gaaaaggag ggtacagaa	420
cagagcaaaa tcagagttca attcagaagc cataagggaag caggggaggg agccatgctt	480
accaacagca aagagtgaag tcaggagtgc agaaaagcaa acaccgtagc	530

<210> 62  
 <211> 448  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 315

<223> 99-25972-317 : polymorphic base C or T

<220>

<221> misc\_binding

<222> 303..327

<223> 99-25972-317.probe

<220>

<221> primer\_bind

<222> 296..314

<223> 99-25972-317.mis

<220>

<221> primer\_bind

<222> 316..334

<223> 99-25972-317.mis complement

<220>

<221> primer\_bind

<222> 1..21

<223> 99-25972.pu

<220>

<221> primer\_bind

<222> 428..448

<223> 99-25972.rp complement

<400> 62

gacctatgag	ttcattggtt	cattccactt	aatctagaga	attcatttta	cctagacaga	60
cactaaaaaa	gacacacagg	tgacacaca	cttaggcata	aattgtcagt	ttttagccca	120
ggtctgtcct	attcccatata	ttacatataa	ttggattaga	aatgtcattt	taaaattatt	180
aagctatagg	taaggccttt	ctttgctaca	tatatatata	atcttgactt	ctttaaaatc	240
tagcctagaa	aacatttttg	tttcttttct	tcttttctct	ggtcccttct	ttagctocca	300
agaatttcag	aaaaytggtg	catttttagat	tatttttgcca	gctgcatact	tttctgtcta	360
atttgcagat	gtcacttact	tgataacact	tcacacagct	gtctaaccoc	agcttaaatg	420
attttctgat	ttctgaagaa	gtggctac				448

<210> 63

<211> 544

<212> DNA

<213> Homo sapiens

<220>

<221> allele

<222> 143

<223> 99-25974-143 : polymorphic base A or G

<220>

<221> misc\_binding

<222> 131..155

<223> 99-25974-143.probe

<220>

<221> primer\_bind

<222> 124..142

<223> 99-25974-143.mis

<220>

<221> primer\_bind

<222> 144..162

<223> 99-25974-143.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..21  
 <223> 99-25974.pu

<220>  
 <221> primer\_bind  
 <222> 524..544  
 <223> 99-25974.rp complement

<400> 63  
 gctataatct ctcccttcctt ggtagccttg ggggtgttta ttttcttagg gttttaagtt 60  
 atttttcttg gcctgaaaac toatggtgcc tactgtcaga gctgcacaga acccagtgtc 120  
 cagccactgc gtgcctccaa gcrctgacat cctaacaaca gtctgggtct ctgaggtctt 180  
 ggagcagagg gaggtgagac aaagataaca ggggtctaac tgtgcctact ggaagtctct 240  
 gctctcaaat aataaaaaaa cttaaaatga acaacatttt actaaatata gcctgatcac 300  
 tgcatttgca tgggggctgg attgatctgg aatgatcaga gctaacaatga agctcttttc 360  
 tggagatgtg ggacatcaaa gtcacatact aagaagggat gaagacattg cagtgtatcg 420  
 gtctccctca ccctgtcctt ctctagcgtt accccctctc tgaatctttt cttactggcc 480  
 acaaagaaca aaactctcag tgcaatgcc a tgacaaattt ctccaggaag sytaaccaa 540  
 tcgt 544

<210> 64  
 <211> 540  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 311  
 <223> 99-25977-311 : polymorphic base A or G

<220>  
 <221> misc\_binding  
 <222> 299..323  
 <223> 99-25977-311.probe

<220>  
 <221> primer\_bind  
 <222> 292..310  
 <223> 99-25977-311.mis

<220>  
 <221> primer\_bind  
 <222> 312..330  
 <223> 99-25977-311.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..21  
 <223> 99-25977.pu

<220>  
 <221> primer\_bind  
 <222> 520..540  
 <223> 99-25977.rp complement

<400> 64  
 caattctctc aaaattgtct caaataaaca agtattgagc catcgtgatg aattatgttt 60  
 gttcacagggt tattccattt aattttctct tcccacctca ttaagatgta aatagtaact 120

```

cagcttcata gagaagactg aggctaagag aaaaggattg tccaggctga aaggttagaa 180
aaaggcagaa gctgctcttc tactcagatt tctgacttca attccacaaa ggtcttaaaa 240
ttaggaaagc aacacttaag cactttaaat agctatgaaa aatgcaaaat ggaatctatt 300
caagcccatg rttctgtgtt aattgtgctc tgatgattgt tctctttctt tgcagcccca 360
cggtaacgga ggtggaagtc agttctactt ctctgtgtgt ttgcaaggaa tgttcctagg 420
aatattctgt tgcaccctgg gaacattatc ttcacccgta ttttaagctt tttatagaca 480
agcaaaaagaa ccatttaaaa tatatatacc acagcttctg atacagtgtt gaaatactac 540

```

&lt;210&gt; 65

&lt;211&gt; 538

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 93

&lt;223&gt; 99-25979-93 : polymorphic base A or G

&lt;220&gt;

&lt;221&gt; misc\_binding

&lt;222&gt; 81..105

&lt;223&gt; 99-25979-93.probe

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 74..92

&lt;223&gt; 99-25979-93.mis

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 94..112

&lt;223&gt; 99-25979-93.mis complement

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 1..19

&lt;223&gt; 99-25979.pu

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 518..538

&lt;223&gt; 99-25979.rp complement

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; 412,462

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 65

```

cactacagta atcgagagaca atgcagtgc aggattgact acccgtggat cagcctgaag 60
cttcaccttt tttttcaaag tgcattttga gartatttgt gctatccatt agtaaaaaatt 120
accaaagagg agcataatgg aagcatagag aaatagtaaa aatagaaaca tccagatatt 180
ctatgaaccc taggaaattt acacaacagt attattattc aatatattgt tgtcgtgtgg 240
caccaacgcc agagaagaat taattgctga gcaacggccc agacaggagt toccaagaag 300
taggttcagg tccatgagtt gggaacttgc tgagttgctg tggatctttg cttctttaaa 360
gtgttgcca tggattgcgg gcatcatcag tattacctgg gatcttcgta gnaatcacag 420
aatcttggtc cctaaccag atctgcagaa tcagttactg cnggcttcac atgataccca 480
ctggatgtgc atgtacattc aagtttgaaa tgatggtgtg aatcaggcag cagaaatg 538

```

&lt;210&gt; 66

&lt;211&gt; 469

<212> DNA  
<213> Homo sapiens

<220>  
<221> allele  
<222> 194  
<223> 99-25985-194 : polymorphic base C or T

<220>  
<221> misc\_binding  
<222> 182..206  
<223> 99-25985-194.probe

<220>  
<221> primer\_bind  
<222> 175..193  
<223> 99-25985-194.mis

<220>  
<221> primer\_bind  
<222> 195..213  
<223> 99-25985-194.mis complement

<220>  
<221> primer\_bind  
<222> 1..21  
<223> 99-25985.pu

<220>  
<221> primer\_bind  
<222> 449..469  
<223> 99-25985.rp complement

<400> 66						
ctttgcttac	aatcaaata	gatacaaggt	atttaaaca	ctaactata	tgtttggaca	60
tctgtgcttt	taaaaataga	tagactagtt	caaaaaaaca	caaataaca	aattcataat	120
aagtaaacac	tttgtcactt	tgtaagtgt	tcagtatcta	taacctactt	aggtattatt	180
tcagcataat	atayattatt	acattaacga	caaaaaaata	aataccttag	atctgtacaa	240
aagtatagac	tgaactagat	aagacaagag	aacattttta	tagatttgaa	ctcagagagc	300
tctgcacgtc	actttaagac	atggctgaac	tgtccaagaa	gcactctgtga	ttattgggggt	360
gtctccctcg	tgaactta	aacatattaa	tctctgacaa	tcacttgcaa	tcaggtaaga	420
ttttcatcct	gagctaattgt	atgctgtgct	gaaaacttta	ctccccatt		469

<210> 67  
<211> 553  
<212> DNA  
<213> Homo sapiens

<220>  
<221> allele  
<222> 398  
<223> 99-25989-398 : polymorphic base A or G

<220>  
<221> misc\_binding  
<222> 386..410  
<223> 99-25989-398.probe

<220>  
<221> primer\_bind  
<222> 379..397

<223> 99-25989-398.mis

<220>

<221> primer\_bind

<222> 399..417

<223> 99-25989-398.mis complement

<220>

<221> primer\_bind

<222> 1..19

<223> 99-25989.pu

<220>

<221> primer\_bind

<222> 533..553

<223> 99-25989.rp complement

<400> 67

cccatcttat	cttcaaattcc	ccatgttaca	ctataggtgc	agtaagtata	aaacaaacct	60
aattagtatt	cttagataat	gcagacattt	aaaaagcaca	agtcattttt	ggaaaacgaa	120
gaagtgcatt	gcoactgagt	ttaccctgaa	atcatgaaaa	tggtctgatg	gagtgcta	180
gtacatatgc	ttacaataaa	aaggtaaaat	ctctcaccat	cacttttggt	agtcatttct	240
ctccgacccc	cacatgtaaa	ctaattccag	ggtaattgca	gggtgatagt	tttgattttc	300
tgtattccaa	gacagtaatt	cttgagagct	atttgcattca	caactactat	cagtggctac	360
taactcaa	aacagcgata	attataataa	tcattacrtta	tttactgagt	attcactatg	420
gactttgcta	accataactca	agcataactgt	ctttccaaga	tacaaatggt	ttagaaaatg	480
aaattttctag	aggtaaggcc	ctggcctcat	tcagattatt	tcagttcata	ttgggattkr	540
aaaacaactg	tga					553

<210> 68

<211> 451

<212> DNA

<213> Homo sapiens

<220>

<221> allele

<222> 264

<223> 99-26146-264 : polymorphic base G or T

<220>

<221> misc\_binding

<222> 252..276

<223> 99-26146-264.probe

<220>

<221> primer\_bind

<222> 245..263

<223> 99-26146-264.mis

<220>

<221> primer\_bind

<222> 265..283

<223> 99-26146-264.mis complement

<220>

<221> primer\_bind

<222> 1..19

<223> 99-26146.pu

<220>

<221> primer\_bind



&lt;222&gt; 436..456

&lt;223&gt; 99-26146.rp complement

&lt;400&gt; 68

ttttacacc	atcaaaacct	aaaaatcctt	catacatctt	aatatctacc	caggaaatga	60
gtttaagtga	atcaattaag	aggaaatgaa	tactcgactt	attcagaccg	aagttgttat	120
ttgggatttg	ttactagtat	gtgtattttt	attatgacac	aactaatgca	gtgatgtgtt	180
atgaaacaaa	tctgagccaa	aaagcagggt	ttgaaatcaa	cttcacttat	aaaggcatta	240
tcttggttac	attatttaaat	atckttttaca	tctacagatg	cttatcacct	taagatggga	300
ttacttcctt	ataattacac	atcataaatt	caaactattg	taagttgaaa	tgcatttagt	360
accttagtaa	acctatttta	atatcaaaaa	aatcccaaat	tgaactatct	taagttgggg	420
accatctgta	gttyyttgca	tgaaaaatag	g			451

&lt;210&gt; 69

&lt;211&gt; 454

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 276

&lt;223&gt; 99-26150-276 : polymorphic base A or G

&lt;220&gt;

&lt;221&gt; misc\_binding

&lt;222&gt; 264..288

&lt;223&gt; 99-26150-276.probe

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 257..275

&lt;223&gt; 99-26150-276.mis

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 277..295

&lt;223&gt; 99-26150-276.mis complement

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 1..19

&lt;223&gt; 99-26150.pu

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 437..454

&lt;223&gt; 99-26150.rp complement

&lt;400&gt; 69

gaaagattag	ttgcatgaga	taatctgac	tctaaaattc	tgccctttta	atattttatc	60
aagggatcca	agaacagtta	ctgcaatatt	cgtaccacaa	tcaactttgc	aatttgaaaa	120
gacaccagct	tggttgagga	gaatgaattg	cgagcggaga	gagtcagaga	ggagataaga	180
ctcagtttga	ggaaggccta	ttgagacatt	tggttcattg	gacaatctgg	tggtttgtac	240
tggtggtcatt	tccacggaga	tgagataaaa	tgagcrtatt	ttggacagtt	tatttaacag	300
catactaact	tttcatattt	tgaggtttct	aaagttgaac	gcattcttcta	attgatgaca	360
tgtcagtttg	atacattggt	tttctctggt	aaaacatat	gctatgatac	atggcatctg	420
agatgcaaag	aaatatggta	cttttggtgg	caca			454

&lt;210&gt; 70

&lt;211&gt; 449

&lt;212&gt; DNA

<213> Homo sapiens

<220>

<221> allele

<222> 44

<223> 99-26153-44 : polymorphic base A or C

<220>

<221> misc\_binding

<222> 32..56

<223> 99-26153-44.probe

<220>

<221> primer\_bind

<222> 25..43

<223> 99-26153-44.mis

<220>

<221> primer\_bind

<222> 45..63

<223> 99-26153-44.mis complement

<220>

<221> primer\_bind

<222> 1..19

<223> 99-26153.pu

<220>

<221> primer\_bind

<222> 429..449

<223> 99-26153.rp complement

<400> 70

atgcttaatg	cacatctacr	taatctatat	aaatcaatca	atcmatctat	gtctacctgt	60
tcagggtata	aaatctttgc	atactacggt	aacatagtat	acaagttcct	tgtgcaacag	120
atagtgata	gcaattaact	gctgtcaagt	tttatctttc	ctgtaaaact	cccagcacct	180
gcactgaact	tcgctggtca	ataagtgcac	tatatacaca	cccatatgca	catacatwat	240
ataatatttt	tatatgtcaa	ttgctcctct	agattgagtc	caatgctcag	aattataacg	300
gtaatttctc	aaaaacaaag	aaacggaagg	aaagaaggaa	gggggatgga	gagaagtcag	360
aaacagagaa	tgagggaaga	agaaaggaag	tgctgaaaat	gtatttgggg	tgaataaaca	420
catgagatgg	tataagagg	caggtaatg				449

<210> 71

<211> 511

<212> DNA

<213> Homo sapiens

<220>

<221> allele

<222> 290

<223> 99-26156-290 : polymorphic base A or C

<220>

<221> misc\_binding

<222> 278..302

<223> 99-26156-290.probe

<220>

<221> primer\_bind

<222> 271..289

<223> 99-26156-290.mis

<220>  
 <221> primer\_bind  
 <222> 291..309  
 <223> 99-26156-290.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..18  
 <223> 99-26156.pu

<220>  
 <221> primer\_bind  
 <222> 491..511  
 <223> 99-26156.rp complement

<400> 71  
 gatgagtgtg aggtagtgtc tcatttttgt tttgagtttt cctaattgatt actgacattg 60  
 aaaatctttt tataatgtct atatttataa aattatattt atatagacaa agaccccaaa 120  
 aaacatattt gtctagggcc ttgtagaccc aaggggaagc cctgatggat tcgaatccta 180  
 attccaccac ctattagctg tgtagtgttg aatatatctg taaacctact tgagtatcag 240  
 aattttttata ttggctgcat tttaactgta gatagctgtt ctggagcaam aatactgtaa 300  
 aattagtctt ttcttgaaaa acaggcagaa cctagtattg acagatataa ctattttttg 360  
 caccctctt ctatatctta tgctattgtc caccagctta ttatattgta acttgatatt 420  
 aattttaaga aagatttgaa attagccata ttttagtatg tcagttttcc ataacaagga 480  
 aaaaaaaca cagaataaag atttaagcca c 511

<210> 72  
 <211> 551  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 189  
 <223> 99-27345-189 : polymorphic base G or C

<220>  
 <221> misc\_binding  
 <222> 177..201  
 <223> 99-27345-189.probe

<220>  
 <221> primer\_bind  
 <222> 170..188  
 <223> 99-27345-189.mis

<220>  
 <221> primer\_bind  
 <222> 190..208  
 <223> 99-27345-189.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..18  
 <223> 99-27345.pu

<220>  
 <221> primer\_bind  
 <222> 531..551  
 <223> 99-27345.rp complement

&lt;400&gt; 72

taacccacac	ggatttagct	agagttctag	ttagctcatg	agtagggaca	gaactggggt	60
tggagttaat	tttactctgt	acttatgact	taaaaccgga	ttggaattac	cttctgagta	120
agtgttccca	tattaccata	tctaaattag	gtgtttcatt	ttttataact	acggtagcat	180
cttctcagsa	tcaactatat	aatttttagg	gccagtgaa	aactgaaact	atgatgtcct	240
tcactagaaa	atgattcaga	attttcagga	ggtgccagca	gagtgttagt	ccaaaagcag	300
gtccttctac	aggaccacga	atgatggcac	agcctgggaa	acaataaagc	tggccctgcc	360
cccttctcga	gtatttatca	gaactttagt	ttgtaatttt	tataaaatac	gtacggtaat	420
ttttctgttg	catttgcttg	ttttctggaa	aactgcaagc	ttcacaacgt	gagtgcacatg	480
catactgtat	ttcttgggtc	agctcagcat	ctagcacagt	gcctgatgca	gcagmcaccg	540
ataaattctg	t					551

&lt;210&gt; 73

&lt;211&gt; 514

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 380

&lt;223&gt; 99-30329-380 : polymorphic base C or T

&lt;220&gt;

&lt;221&gt; misc\_binding

&lt;222&gt; 368..392

&lt;223&gt; 99-30329-380.probe

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 361..379

&lt;223&gt; 99-30329-380.mis

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 381..399

&lt;223&gt; 99-30329-380.mis complement

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 1..18

&lt;223&gt; 99-30329.pu

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 496..514

&lt;223&gt; 99-30329.rp complement

&lt;400&gt; 73

gacatacccc	aaacacagca	gcacaaagca	actgcctgtg	attgtcaatt	atctgcacta	60
actttgaata	gacaactgag	agtaaaattg	accattagct	gttattaatc	ttttaaatag	120
cttattttatc	aggacaatat	tttctttgag	gaatggtttc	acatcgttga	taaacagttt	180
cctattacta	acattaaact	aaataatata	tgcatatatt	cttttttacc	cttagaaagt	240
taatctaata	atattattct	gattagtggg	acttttttta	atttttgtaa	tggtgtttat	300
ggtgtttgat	ttgtactttt	ttgtcttttt	catacctttt	aataacatcc	gatttactac	360
tggagcattt	tatttttcaty	gtctctttaa	atacttaata	acttttagat	ttgttggtca	420
caaaactaaa	taattgcttg	atgtttcatt	ttatcttctt	tattaaatag	gcagctcgtg	480
tacatatgaa	agtgtctttt	tatacaccca	aacc			514

&lt;210&gt; 74

&lt;211&gt; 511

<212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 342  
 <223> 99-31798-344 : variable motifCAT or ACTGTCTCTACCTCA

<220>  
 <221> primer\_bind  
 <222> 1..20  
 <223> 99-31798.pu

<220>  
 <221> primer\_bind  
 <222> 491..511  
 <223> 99-31798.rp complement

<400> 74  
 caaccataa caaatggaag gcagagccca atggcaaagg tagtgccaag gcagatgtgt 60  
 tccccacagc ctagaagttg gggagagagg aggcctaaaa tgcttccaga ctgggtctac 120  
 catagaggaa ggtaagaagg taagatacct taagattcgg gcattgggaa ggggcagaga 180  
 aggaaaattg catgatctca tactaggtga cctcagtacc agtgaagttt tgattggatt 240  
 tgcgagctga tctaagccaa ctccaacatt caacaatggt aagggaaccc ttagraattc 300  
 acttgtaggt aagggtaca tatattcttc attacaacct gactgtctct acctcattat 360  
 gcatttatat ccttattata gcagaggtag ttatgattaa agtaactatc agactgaaat 420  
 cctcccaaag ggtaaacatt attagcaagc aggaaggaaa caagatatgt ctaggagaga 480  
 tttttaagtg cagaaatatc tccactccgt c 511

<210> 75  
 <211> 474  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> allele  
 <222> 159  
 <223> 99-5873-159 : polymorphic base C or T

<220>  
 <221> misc\_binding  
 <222> 147..171  
 <223> 99-5873-159.probe

<220>  
 <221> primer\_bind  
 <222> 140..158  
 <223> 99-5873-159.mis

<220>  
 <221> primer\_bind  
 <222> 160..178  
 <223> 99-5873-159.mis complement

<220>  
 <221> primer\_bind  
 <222> 11..28  
 <223> 99-5873.pu

<220>  
 <221> primer\_bind

&lt;222&gt; 466..474

&lt;223&gt; 99-5873.rp complement

&lt;400&gt; 75

aaaggagcag gcgtaacaat aagcaggggt agtcgccaca aaacttgaga taagaggaaa	60
actaaaaaag tctaataaaa tcagtagtct taaaaagatg acatgatagg aagagaagtg	120
ttaaaaaaga aaaaaaatag gtatgaaaga gagtaacaya taccggaaaa gggataaaat	180
acatcctttg aaagaacaaa gagttattca aattgaattc ttaatgaatt acttaaacag	240
cagattagat attgttaaaa agaggaatag ggaattaaat gatatatgtg atgatattac	300
ctagtgtaac catcaaagat gtattgcaaa tgataaagaa aaaaatgctg ccatggcaat	360
attaatatca taaaaatata ctttaagaag taaataaatg caactaggaa tagagaaagc	420
taatgaataa taatatitaa aaaaaagtat aacaagtata cataagatgt aata	474

&lt;210&gt; 76

&lt;211&gt; 490

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; allele

&lt;222&gt; 220

&lt;223&gt; 99-6012-220 : polymorphic base G or T

&lt;220&gt;

&lt;221&gt; misc\_binding

&lt;222&gt; 208..232

&lt;223&gt; 99-6012-220.probe

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 201..219

&lt;223&gt; 99-6012-220.mis

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 221..239

&lt;223&gt; 99-6012-220.mis complement

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 11..29

&lt;223&gt; 99-6012.pu

&lt;220&gt;

&lt;221&gt; primer\_bind

&lt;222&gt; 477..490

&lt;223&gt; 99-6012.rp complement

&lt;400&gt; 76

attgtattta gtcttgactt gttttcctga ggggccaggt tgatttgcac gctcttgagg	60
aaatatacac gtcttctcag ttttaataat tgactgacag ccctgtgggt tctcaggacc	120
cagtgaagctg ctgctcccag gtcagctctgc aaaggatgct gggtcccttg tggctctcatc	180
aagggtgagga atttctctgat ttttagagatt tctttatcck aattttgaag actttctttc	240
acatttctag gcataaaaaa atgtacagca ctctactgct tgtttaacaa atggatagtg	300
atatactctgc caacaaagac cacatggagt atttcattga ctatcagaga agtttctctg	360
aaaggcaccac tacttagtgt tttatttcca tgagtgaagg aaaattagtt atttgaagta	420
tttggtgtgc ttttagttgtt tctaaagtag tgctgatttt atatgcccat aatattcata	480
tatacaccca	490

&lt;210&gt; 77

&lt;211&gt; 529

<212> DNA  
<213> Homo sapiens

<220>  
<221> allele  
<222> 99  
<223> 99-6080-99 : polymorphic base C or T

<220>  
<221> misc\_binding  
<222> 87..111  
<223> 99-6080-99.probe

<220>  
<221> primer\_bind  
<222> 80..98  
<223> 99-6080-99.mis

<220>  
<221> primer\_bind  
<222> 100..118  
<223> 99-6080-99.mis complement

<220>  
<221> primer\_bind  
<222> 11..28  
<223> 99-6080.pu

<220>  
<221> primer\_bind  
<222> 519..529  
<223> 99-6080.rp complement

<400> 77  
acaacaacaa aaatgtgtcc ctgaaaccca tgctatatcc aactgaatat tctaattgtct 60  
ttgattacaa agccatctct agcaatttaa tacaattayg aaatggaaaa gttggcaaat 120  
gcaaaacaat agctcgtgtt caaggtatgt ctttattagg ggaagtttat cgaaacagat 180  
gtttatgcta ttccctataa actagattct aaaatatttt attctataaa gatgtattga 240  
ctttatatga aaaaattatt gaaaaatcta caagatgggtg aaactcttta gaactatatt 300  
tctattacaa gtttattttt aatttcacaa atgtactgca taaatgcagc aaaaccttta 360  
ttgtcacata ttaaaacatg tacattattg tgtgcaaatt aaaatttcac taccttaaac 420  
caaaaagtga gttggccaga tagtaataaa tttaggctct aaggctgaaa agcgcttgta 480  
ttaattactc aactccacca ctattttgcc aaagcagtca cagacaata 529

<210> 78  
<211> 489  
<212> DNA  
<213> Homo sapiens

<220>  
<221> allele  
<222> 156  
<223> 99-7308-157 : polymorphic base C or T

<220>  
<221> misc\_binding  
<222> 144..168  
<223> 99-7308-157.probe

<220>  
<221> primer\_bind

<222> 137..155  
 <223> 99-7308-157.mis

<220>  
 <221> primer\_bind  
 <222> 157..175  
 <223> 99-7308-157.mis complement

<220>  
 <221> primer\_bind  
 <222> 1..18  
 <223> 99-7308.pu

<220>  
 <221> primer\_bind  
 <222> 469..489  
 <223> 99-7308.rp complement

<400> 78  
 tgtggtctgg atatggtgra ctgtccttca cacacagatg tgggaagcca tgatcatcag 60  
 ttgcattatt cctgaggggc aatgcattcc agttacatag aaccagtttc tacgtttcag 120  
 ggatatgta ttcatggtga caaatttatt cacatyttaa gtaattttaa gtaattcaca 180  
 ttttaagtaa ttttcttgaa tgtgcctcat tggcttctgt gcctcttcag aaaagatgaa 240  
 ctaaacactg gcataatgtg tcagatttca acattccgtt gttttcattg tggataattt 300  
 ctgtcccata tttttgtgta aagtttagaca ataaagtgtt aatattctgg cgtcggcaca 360  
 ttttctttcc tgataaataa caattcacat atctttttaa aatatcagag aatatagtaa 420  
 ccaatttoca attctttttt caccatgtat ctattggagt tttaaaatga ctaatactaa 480  
 ggcaactat 489

<210> 79  
 <211> 112  
 <212> DNA  
 <213> Pan troglodytes

<400> 79  
 tcccagtggg attcaaccac tcatagactg cttttgttca aactgcgttc aaataaggca 60  
 aacgccgagc tggaaccaat ctcactgttt ctgtacctcg cttctaattc ct 112

<210> 80  
 <211> 118  
 <212> DNA  
 <213> Pan troglodytes

<400> 80  
 tttcagaaaa caaaaactcc taagagaatg aatagaaatt atcttaagga attataaaga 60  
 accatctcag aattttactc tgttactgca aaatctcatg caaagataac caagtatg 118

<210> 81  
 <211> 146  
 <212> DNA  
 <213> Pan troglodytes

<400> 81  
 ttgcaggctc acattcggtc tagaggctgc ctgcggctct tagaggccac tcatagttcc 60  
 ttgccatgaa gacttttcca acatggctgc ttacttctta aagcctgtaa ggagatccag 120  
 tctcagttgg ccaagtcttg taaaag 146

<210> 82  
 <211> 94  
 <212> DNA  
 <213> Pan troglodytes



<400> 82  
gaaaaataacg catttataaa acccagctca gatgttactt ccaaccaaca aatttcctgg 60  
acacactaag taatagagtt aatcacactg tcct 94

<210> 83  
<211> 229  
<212> DNA  
<213> Pan troglodytes

<400> 83  
ttgcagggtcc acattcgttc tagaggctgc ctgcgggtcct tagaggccac tcatagttcc 60  
ttgccatgaa gacttttcca acatggctgc ttacttctta aagcctgtaa ggagatccag 120  
tctcagttgg ccaagtcttg taaaaggtaa cctcatcacc ggagtcaagt ctcacacct 180  
ttcccatggt ccattggtta gaagtcaccg ctctgcccca cattcaagg 229

<210> 84  
<211> 205  
<212> DNA  
<213> Pan troglodytes

<400> 84  
tgaaggtctc cctgctgtgc ttagagaggg ggatgctgga agaagagtgg tcagagaagc 60  
aacattgctg gctctaaaga cagaggacag gccacgagcc aaggcatgca ggaggcctcc 120  
agaagctggg gcaggcaaga gaatggactc agccctggag cctccagaag gcacgcagcc 180  
ctgccattc cttgctttta gccag 205

<210> 85  
<211> 82  
<212> DNA  
<213> Pan troglodytes

<400> 85  
aaacaatacg aacagaaaca aagctatgta cttgttttgg gaagcaacct caaatggaaa 60  
gatgggggct gtgttctaaa ag 82

<210> 86  
<211> 234  
<212> DNA  
<213> Pan troglodytes

<400> 86  
gaacattcac tctgggagct ctgtgctgcc atgcaaatgt cacattgcac tgaggccgcc 60  
atgctgtaaa aaacccaag ctacctacct gaaaagctgg agagaatcag caccatcag 120  
cttttcagc attttgggtc ctcccagcta aatcatcggc attgtgaagc agagatgaga 180  
catcctgcca tgccctttcc aaatcctgga ccaacagaat catcatatac acta 234

<210> 87  
<211> 714  
<212> DNA  
<213> Pan troglodytes

<400> 87  
gaacattcac tctgggagct ctgtgctgcc atgcaaatgt cacattgcac tgaggccgcc 60  
atgctgtaaa aaacccaag ctacctacct gaaaagctgg agagaatcag caccatcag 120  
cttttcagc attttgggtc ctcccagcta aatcatcggc attgtgaagc agagatgaga 180  
catcctgcca tgccctttcc aaatcctgga ccaacagaat catcatatac actagttagc 240  
ttttacttta agccactaat ttgggtgta gctcttttgc agcaaaagat aatcagaaca 300  
ctgcatgca ggagctgaga aacttgetca ctgttatgcg tgtatccaat gactactcac 360  
acacacatta cttgcttcac tctgctttct ggctttccat tcagcaaatt ttaagaaatg 420  
aaaaagttat tcatgggaaa ttaactcaag atgaatattg ttgtagttaa tctctttgta 480

aatggtaaaa	tattaaatta	ttgtgagtca	tttttattta	aagaagtttg	gttcttttaa	540
tgaaaaaagt	ttgggtcttt	tgtggatact	ttgcatatag	acattctaga	catgaaagaa	600
gtttgaagac	aaagaatgtt	gaaggaatat	agacatgaaa	gaatgttgaa	gtttgtttt	660
ttgtggatac	tctgcatata	gacattctag	actggagtga	tacatcttgt	tcag	714

&lt;210&gt; 88

&lt;211&gt; 220

&lt;212&gt; DNA

&lt;213&gt; Pan troglodytes

&lt;400&gt; 88

acatgaaaga	atgttgaagt	ttgggttttt	gtggatactc	tgcatataga	cattctagac	60
tggagtata	catcttggtc	aggtttctat	gttttcaaaa	tgataacttg	acacacctac	120
caatgccaat	ttttaggata	tatatctgta	tatctgtgtg	aactaaggaa	aattcaggta	180
tgattaaacc	tacatatatt	gccaggagag	actggataat			220

&lt;210&gt; 89

&lt;211&gt; 173

&lt;212&gt; DNA

&lt;213&gt; Pan troglodytes

&lt;220&gt;

&lt;221&gt; misc feature

&lt;222&gt; 91..92,131

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 89

ggtcctgata	ttcccagtga	cttagcagca	cagtaatttg	gctgaaatgt	tgatgagttc	60
agacaagaag	aagggtgatga	tgacaagaga	nncgtcacgg	ggaaattctg	ccttgaggaa	120
ggaatagttt	naggctggag	agatgtgcta	cccccgccac	tctttctcaa	atg	173

&lt;210&gt; 90

&lt;211&gt; 1105

&lt;212&gt; DNA

&lt;213&gt; Pan troglodytes

&lt;220&gt;

&lt;221&gt; misc feature

&lt;222&gt; 201,744,803,808..809

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 90

gaaagacaac	ctgtaggcac	agaaatcatt	gctctgctta	caactgaaatc	tcctgctgac	60
tttgagagct	gagcaactgga	ttttgatggg	agagtgaagg	ttagtgaagg	aatgatttga	120
acotttgttg	atctaatact	tctgtgatac	agttatctaa	acagaatgag	tctaaagtca	180
gtataaagac	ctgtgtctct	ncatgttttg	ctacttataa	ttgcagagga	acattttgag	240
atagaaaatt	ttatccccct	ttcaagtgc	tttttataaa	acataacctt	ttaagagagg	300
tcattctttt	gaaaaattgt	ctgaatacaa	gagaagtaac	tgaacagcaa	gggattgtag	360
ccataatgtt	cttgaccact	ttagattgac	aaatagtgtc	tcaaaattga	tcagggttctg	420
gagaatctat	tgaggttgta	tggagtaacc	aagtgatgga	agaggttgga	agtgaataat	480
tttatgtggt	attatgtgat	tttttttggt	gggggtagt	tctggagctg	agcatgatgg	540
ggcctagatg	gggcagcctt	ctaagtctcc	cagtatttctg	caaatcatat	gtcacgtcaa	600
aattgtaatg	acaaagctca	cagcagctgc	ctacgccagg	tttgggtgcca	gctctaatta	660
gagcatcgat	catttctctc	cccttcatga	atgcacgtgc	gtgcacattt	tcatatacaa	720
gggcaaagga	catttcaaa	tttnggatgg	aatctttcct	acttagaagt	aataaaaagc	780
aatactttga	aatcttacca	aanattannc	aaatagcaac	aatttttcca	tttaaagatg	840
tgtgtcttca	ggtctgttgg	atctttcaaa	tttctatttt	ggacatggaa	actacagatg	900
aaatttaact	tactcaaaagt	tacacagaac	tcctggctga	tctcttttac	ttaattcttc	960
taaaatttta	agtaagaata	attttgaat	gttattggta	atgcttaaat	aacaaaagtt	1020
actttaatat	gttacatgtt	ggaagtttca	tttatactta	atatctaatt	taactcttgt	1080
gaatatatta	ttcagttttt	ttttt				1105

<210> 91  
 <211> 114  
 <212> DNA  
 <213> Hylobates sp.

<220>  
 <221> misc\_feature  
 <222> 45,75  
 <223> n=a, g, c or t

<400> 91  
 tcccagtgag attcaaccac tcatagactg cttttgtcca aactngcggt caaataaggc 60  
 aaacgccgag ctggnaaacc atctcactgt ttctgtacct cacttctgat tcct 114

<210> 92  
 <211> 117  
 <212> DNA  
 <213> Hylobates sp.

<400> 92  
 ttttagaaaa caaatctcct aagagaatga atagaaatta tcttaaggaa ttataaagaa 60  
 ccatctcaga attttactct gttactgcaa aatctcatgc aaagatagcc aagtatg 117

<210> 93  
 <211> 146  
 <212> DNA  
 <213> Hylobates sp.

<400> 93  
 ttgcaggctc acattcggtc tagaggctgc ctgcagtcct tagaggccac tgatacttcc 60  
 ttgccatgag gacttttcca acatggctgc ttacttcttc aagcctgtaa ggagatctgg 120  
 tctcagttgg ccaagtcttg tgtaagtaa cctcatcacc ggagtcacgt ctcatcacct 146

<210> 94  
 <211> 229  
 <212> DNA  
 <213> Hylobates sp.

<400> 94  
 ttgcaggctc acattcggtc tagaggctgc ctgcagtcct tagaggccac tgatacttcc 60  
 ttgccatgag gacttttcca acatggctgc ttacttcttc aagcctgtaa ggagatctgg 120  
 tctcagttgg ccaagtcttg tgtaagtaa cctcatcacc ggagtcacgt ctcatcacct 180  
 ttccaagtt coattgggta gaagtcacc ctcctgccca cattcaagg 229

<210> 95  
 <211> 205  
 <212> DNA  
 <213> Hylobates sp.

<220>  
 <221> misc\_feature  
 <222> 187  
 <223> n=a, g, c or t

<400> 95  
 tgaagttctc cctgctgtgc tcagagaggg ggggtgctga agaagagtgg tcagagaagc 60  
 aacattgctg gctctaaaga gagaggacag gccacgagcc aaggcatgca ggaggcctcc 120  
 agaagctggg gcaggcaaga gaatggactc ggccctgcag cctccagaag gcatgcagcc 180  
 ctgcccnttc cttgctttta gccag 205

<210> 96  
 <211> 234  
 <212> DNA  
 <213> Hylobates sp.

<400> 96  
 gaacattcac tctgggagct ctgtgctgcc atgcaaagt cagattgcac tgaggccacc 60  
 atgtgtgaag aaacctcaag ctaccaacct gaaaagctgg agagaatcag caccatcag 120  
 cttttccagc cttttgggtc ctcccggcta aatcataggc attgtgaagc agagatgaga 180  
 cctcctgcca tgccctttcc aaatcctgga ccaacagaat catcatatac aata 234

<210> 97  
 <211> 715  
 <212> DNA  
 <213> Hylobates sp.

<400> 97  
 gaacattcac tctgggagct ctgtgctgcc atgcaaagt cagattgcac tgaggccacc 60  
 atgtgtgaag aaacctcaag ctaccaacct gaaaagctgg agagaatcag caccatcag 120  
 cttttccagc cttttgggtc ctcccggcta aatcataggc attgtgaagc agagatgaga 180  
 cctcctgcca tgccctttcc aaatcctgga ccaacagaat catcatatac aataggaagc 240  
 ttttacttta agccactaat ttgggggtgta gttcttatgc agcaaaagat aatcagaaca 300  
 ctgccatgca ggagctgata aacttgctta ctattatgtg tatatccaat gactactcac 360  
 acacacattt cttgcttcac tctactttct ggctttccat tcagcaaatt ttaagaaatg 420  
 aaaatgttat tcatgggaaa ttaactcaag atgtatatgt ttgtagttaa actctttgta 480  
 aatggtaaaa tattaataa ttgctaata tctttattta aagaagtttg gttctttaaa 540  
 tgaaagaagt ttgggtcttt tgtggatact ctgcatatag acattctaga catgaaagaa 600  
 tgttgaagac aaagaatgtg gaaggcatat agacatgaaa gaatgttgaa gtttggttct 660  
 tttgtggata ctctgcatat agacattcta gactggagt atatatcttg ttcag 715

<210> 98  
 <211> 213  
 <212> DNA  
 <213> Hylobates sp.

<400> 98  
 acatgaaaga atgttgaagt ttggttcttt tgtggatact ctgcatatag acattctaga 60  
 ctggagtgat atatcttgtt caggtttcta tgtttttaaa atgataactt gacacaccta 120  
 ccaatgcaaa tttttaggat atatatctgt gtgaactaag gaaaattcag gaatgattaa 180  
 acctacatat tttctcagga gagactggat aat 213

<210> 99  
 <211> 209  
 <212> DNA  
 <213> Hylobates sp.

<220>  
 <221> misc feature  
 <222> 26..27  
 <223> n=a, g, c or t

<400> 99  
 ggttttgtgc atttctaaga tgatcnctt ggtgtggggt cctgatcttc ccagtgaactt 60  
 agcagcacag taatttggct gaaatgttga tgagttaaga caagaagaag gtgacgctga 120  
 caagagagcg tcaaggggaa attctgccct gaggaaggaa tagtttgagg ctggagagat 180  
 gtgctacccc cgccactctt tctcaaatg 209

<210> 100  
 <211> 1415  
 <212> DNA  
 <213> Hylobates sp.

&lt;220&gt;

&lt;221&gt; misc feature

&lt;222&gt; 343,383,387,570,1118

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 100

```

gaaagaaaac ctgtaggcac agaaatcatt gctctgctta cactgaaatc tctgctgac      60
tttgagaggt gagtggttga ttttgatgga aggggtgaaac ttagtgaaga aatgatttga      120
accttggttg atcgaatact tctgtgatac agttatctaa acagaatgag tctaagggtca      180
gtataaagac ctgtgtctct tcgtgttttg ctactataaa ttgcagagga acgttttgag      240
atagaaaatt ttatccact ttcaagtgaac tttttataaa acataacctt ttaagagagt      300
tcattctttt gaaaaattgt ctgaatacaa gaaaagtaac tgnacagca aggattgtag      360
ccataatgtt cttgaocact ttnagantga caaatagtgt ctcaaaattg atcagggtgcc      420
acagaatcta tagaggttgt atggagtaac caactgatgg aagaggttgg acacatctga      480
ataattttat gtggtattat gtgatgttat tttggtgggg tagtgtctgg agctgagcat      540
gatggggcct agatggggca gccttctaan gctcccaaca tttcccaaat catatgtcat      600
ctcaaaattg taatgacaaa gctcacagca gctgcctaca ccaggtttgg tgccagctct      660
aattagagca tcgagcattt ctctcccctt catgaatgca catgctgca cattttcata      720
tatgagggca aaggacattt caaagtgttg ttggaatctt tcctacttag aagtaataaa      780
aacaatactt tgaaatctta ccaaataa aacaaatagc aacaattttt ccatttaaag      840
gtgtgtgtct tcagggtctgt tggatctgtc aaatttctat tttggacatg gaaaccacag      900
atgaaattta acttacacaa agttacacag aactcctggc tgatctctt tacttaattc      960
ttctaaaacc taaagtaaga ataattttga aattttattg gtaatgctta aataacagtt     1020
actttaatat attacatgtt tgaagatttt catttatact taatatcgaa ttttaacttgt     1080
gaaaatatta ttcagtattt tgttttgttt tgttttcnaa aatgacttgt tcagggtccc     1140
ttttgggatt ttaagtgaga aaaactctat ttcataatat tgttaccatg ctcatccaat     1200
aacacttcat agatgggttag ccactcttaa aattaaattt ttagtttgag ggccattgaa     1260
aataccacct atattgggat ttctgtttta cacttcaaaa tgtatcatta tcataattta     1320
aaattagatt taaatacaag cctctttata gaaaacctga taaatatggc tttatttgtt     1380
tgcttggttt ttttgcaggg agggatattc ccttg                                     1415

```

&lt;210&gt; 101

&lt;211&gt; 112

&lt;212&gt; DNA

&lt;213&gt; Gorilla gorilla

&lt;220&gt;

&lt;221&gt; misc feature

&lt;222&gt; 99

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 101

```

tcccagtagg attcaaccac tcatagactg cttttgttca aactgcgttc aaataaggca      60
aacgccgagc tggaaccaat ctcaactgtt ctgtacctng cttctaattc ct                112

```

&lt;210&gt; 102

&lt;211&gt; 118

&lt;212&gt; DNA

&lt;213&gt; Gorilla gorilla

&lt;400&gt; 102

```

tttcagaaaa caaaaactcc taagagaatg aatagaaatt atcttaagga gttataaaga      60
accatctcag aattttactc tcttactgca aaatctcatg caagataac caagtacg        118

```

&lt;210&gt; 103

&lt;211&gt; 146

&lt;212&gt; DNA

&lt;213&gt; Gorilla gorilla

&lt;400&gt; 103

ttgcagggtcc acattcggttc tagagggtgc ctgcggtcct tagaggccac tcatagttcc 60  
 ttgccatgaa gacttttcca acatgactgc ttactttctc aagcctgtaa ggagatccag 120  
 tctcagtcgg ccaagtcttg taaaag 146

<210> 104

<211> 94

<212> DNA

<213> Gorilla gorilla

<400> 104

gaaaataacg catttataaa acccagctca gatgttactt ccaaccaaca aatttcctgg 60  
 acacactaag taatagagtt aatcacaccg tcct 94

<210> 105

<211> 229

<212> DNA

<213> Gorilla gorilla

<400> 105

ttgcagggtcc acattcggttc tagagggtgc ctgcggtcct tagaggccac tcatagttcc 60  
 ttgccatgaa gacttttcca acatgactgc ttactttctc aagcctgtaa ggagatccag 120  
 tctcagtcgg ccaagtcttg taaaaggtaa cctcatcacc ggagtcacgt ctcacacct 180  
 ttcccatgtt ccattgggta gaagtcccag ctcccgcca cattcaagg 229

<210> 106

<211> 205

<212> DNA

<213> Gorilla gorilla

<400> 106

tgaagggtctc cctgctgtgc ttagagaggg ggacgctgga ataagagtgg tcagagaagc 60  
 aacattgctg gctctaaaga cagaggacag gccacgagcc aaggcatgca ggaggcctcc 120  
 agaagctggg gcaggcaaga gaatggactc agccctggag cctccagaag gcacacagcc 180  
 ctgcccgttc cttgctttta gccag 205

<210> 107

<211> 82

<212> DNA

<213> Gorilla gorilla

<400> 107

aaacaatacc aacagaaaca aagctatgta cttgttttgg gaagcaacat caaatggaaa 60  
 gacgggggct gtgttctaaa ag 82

<210> 108

<211> 234

<212> DNA

<213> Gorilla gorilla

<220>

<221> misc\_feature

<222> 192

<223> n=a, g, c or t

<400> 108

gaacattcac tctgggagct ctgtgctgcc atgcaaatgt cacattgcac tgaggccgcc 60  
 atgtgtgaaa aaaccccaag ctacctacct gaaaagctgg agagaatcag caccatcag 120  
 cttttccagc attttgggtt ctcccagcta aatcattggc attgtgaagc agagatgaga 180  
 cctcctgcca tncctttcc aaatcctgga ccaacagaat catcatatac aata 234

<210> 109

<211> 714  
 <212> DNA  
 <213> Gorilla gorilla

<220>  
 <221> misc\_feature  
 <222> 192  
 <223> n=a, g, c or t

<400> 109  
 gaacattcac tctgggagct ctgtgctgcc atgcaaagt caccattgcac tgaggccgcc 60  
 atgctgtaaa aaacccaag ctacctacct gaaaagctgg agagaatcag caccatcag 120  
 cttttccagc attttgggtt ctcccagcta aatcattggc attgtgaagc agagatgaga 180  
 cctcctgcca tncctttcc aaatcctgga ccaacagaat catcatatac aatagtaagc 240  
 ttttacttta agccactaat ttgggggtgta gctcttttgc agcaaaagat aatcagaaca 300  
 ctgccatgca ggagctgata aacttgctca ctgttatgta tgtgtccaat gactactcac 360  
 acacacatta cttgcttcac tctgctttct ggctttccat tcagcaaatt tgaagaaatg 420  
 aaaaagtatt tcatgggaaa ttaactcaag atgaatattg ttgtagttaa tctctttgta 480  
 aatggtaaaa tattaataa ttgtgagtag tttttattta aagaagtttg gttctttaaa 540  
 tgaaagaagt ttgggtcctt tgtggatacg ttgcatatag acattctaga catgaaagaa 600  
 tgttgaagac aaagaatgtt gaaggaatat agacatgaaa gaatgttgaa gtttgtttt 660  
 ttgtggatac tctgcatata gacattctag actggagtgta tacatcttgt tcag 714

<210> 110  
 <211> 220  
 <212> DNA  
 <213> Gorilla gorilla

<220>  
 <221> misc\_feature  
 <222> 122  
 <223> n=a, g, c or t

<400> 110  
 acatgaaaga atgttgaagt ttgttttttt gtggatactc tgcatataga cattctagac 60  
 tggagtgata catcttggtc aggtttctat gtttttaaaa tgataacttg acacacctac 120  
 cnatgccaat ttttagggta tatatctgta tatctatgtg aactaaggaa aattcaggaa 180  
 tgattaaacc tacatatatt gccaggagag accggataat 220

<210> 111  
 <211> 209  
 <212> DNA  
 <213> Gorilla gorilla

<220>  
 <221> misc\_feature  
 <222> 97,115,117,130,134,160,193  
 <223> n=a, g, c or t

<400> 111  
 ggttttgtgc atttctaaga tgatcatctt ggtgtggggt cctgatcttc ccagtgactt 60  
 agcagcacag taatttggct gaaatgttga tgagttnaga caagaagaag gtgangntga 120  
 caagagagcn tcanggggaa attctgccct gaggaaggan tagtttgagg ctggagagat 180  
 gtgctacccc cgncaactctt tctcaaagt 209

<210> 112  
 <211> 1105  
 <212> DNA  
 <213> Gorilla gorilla

<220>

<221> misc\_feature  
 <222> 959  
 <223> n=a, g, c or t

<400> 112  
 gaaagaaaac ctgtaggcac agaaatcatt gctctgctta tactgaaatc tcctgctgac 60  
 tttgagagct gagcattgga ttttgatggg agagtgaac ttagtgaaga aatgatttga 120  
 acctttgttg atctaatact tctgtgatac agttatctaa acagaatgag tctaaagtca 180  
 gtataaagac ctgtgtctct ccatgttttg ctacttataa ttgcagagga acattttgag 240  
 atagaaaaatt ttatccccct ttcaagtgc tttttataaa gcgtaacctt ttaagagagg 300  
 tcattctttt gaaaaattgt ctgaatacaa gagaagtaac tgaacagcaa gggattgtag 360  
 ccataatggt cttgaccact ttagattgac aaatagtgtc tcaaaattga tcaggttccg 420  
 gagaatctat tgagggtgta tggataaacc aagtgtgga agagggtgga agtgaataat 480  
 tttatgtggt attatgtgat gtttttttgg tgggggtagt gtctggagct gagcgtgatg 540  
 ggccctagat ggggcagcct tctaagtctc ccagcatttc gcaaatacata tgtcacgtca 600  
 aaattgtaat gacaaaagctc acagcagctg cctacgccag gtttggtgcc agctctaatt 660  
 agagcatcga tcatttctca cgccttcctg aatgcacatg tgtgcacatt ttcatatata 720  
 agggcaaaag acatttcaaa gtttggtgga aatctttcct acttagaagt aataaaaaac 780  
 aatactttga aatcttacc aatattaaat aaatagcaac aatttttcca tttaaagatg 840  
 tgtgtcttaa ggtctgttgg atctttcaaa tttctatttt ggacatggaa actacagatg 900  
 aaattttaact gacacaaagt tacacagaac tcttggtgta tctgttttac ttaattctnc 960  
 taaaatctaa agtaagaata attttgaaat tttattggtg atgcttaaat aacaaaagtt 1020  
 actttaatat gttacatgtt tgaagtttca tttatactta atatctaatt taactcttgt 1080  
 gaaaatatta ttcagtattt ttttt 1105

<210> 113  
 <211> 112  
 <212> DNA  
 <213> Pongo pygmaeus

<220>  
 <221> misc\_feature  
 <222> 69..71,99,106  
 <223> n=a, g, c or t

<400> 113  
 tcccagtgagg attcaaccac tcatagactg cttttgttcg aactgcgttc aaataaggca 60  
 aacgccgann nggaaccaat ctcaactgtt ctgtacctna cttatnattc ct 112

<210> 114  
 <211> 118  
 <212> DNA  
 <213> Pongo pygmaeus

<400> 114  
 tttcagaaaa caaaaactcc taagagaatg aatagaaatt atcttaagga attataaaga 60  
 accgtctcag aattttcttc tgttactgca aaatctcatg caaagatgac caagtatg 118

<210> 115  
 <211> 146  
 <212> DNA  
 <213> Pongo pygmaeus

<400> 115  
 ttgcaggctc acattcgttc tagaggctgc ctgcagtcct tagaggtcac tcatagttca 60  
 ttgccatgag gacttttcca acatggctgc ttacttcttc aagcctgtaa ggagatctgg 120  
 tctcagttgg ccaaatcttg tataag 146

<210> 116  
 <211> 94  
 <212> DNA



&lt;213&gt; Pongo pygmaeus

&lt;400&gt; 116

ggaaataacg catttataaa acccagctca gatattactt ccaaccaaca aattgcctgg 60  
 acacactaag taatagagtt aatcacacca tctt 94

&lt;210&gt; 117

&lt;211&gt; 229

&lt;212&gt; DNA

&lt;213&gt; Pongo pygmaeus

&lt;400&gt; 117

ttgcaggtcc acattcggtc tagaggctgc ctgcagtcct tagaggtcac tcatagttca 60  
 ttgccatgag gacttttcca acatggctgc ttactttctc aagcctgtaa ggagatctgg 120  
 tctcagttgg ccaaatcttg tataaggtaa cctcatcacc ggagtcattgt ctcacacctc 180  
 ttcccggtt ccattgggta gaagtcaccg ctccctgccca cattcaagg 229

&lt;210&gt; 118

&lt;211&gt; 234

&lt;212&gt; DNA

&lt;213&gt; Pongo pygmaeus

&lt;400&gt; 118

gaacattcac tctgagagct ctgtgctgcc atgcaaatgt cagattgcac taaggccgtc 60  
 atgctgtaag aaaccccaag ctacctacct gaaaagctgg agagaatcag caccatcag 120  
 cttttccagc cctttgggtc ctcccagcta aatcataggt attgtgaagc agagatgaga 180  
 cctcctgcc a tgcctttcc aaatcctgga ccaacagaat catcatatac aata 234

&lt;210&gt; 119

&lt;211&gt; 599

&lt;212&gt; DNA

&lt;213&gt; Pongo pygmaeus

&lt;400&gt; 119

gaacattcac tctgagagct ctgtgctgcc atgcaaatgt cagattgcac taaggccgtc 60  
 atgctgtaag aaaccccaag ctacctacct gaaaagctgg agagaatcag caccatcag 120  
 cttttccagc cctttgggtc ctcccagcta aatcataggt attgtgaagc agagatgaga 180  
 cctcctgcc a tgcctttcc aaatcctgga ccaacagaat catcatatac aatagtaagc 240  
 ttttacttta agccactaat ttgggggtga gttcttttgc agcaaaagat aatcagaaca 300  
 ctgccatgca ggagctgata aacttggtca ctgttatgta tgtatccaat gactactcac 360  
 acacacatta cttgcttcac tctactttct ggctttccat tcagcaaat ttaagaaatg 420  
 aaaatgttat tcatgggaaa ttaactcaag atgaatatgg ttgtagttaa tctctttgtc 480  
 aatggtaaaa tattaatta ttgtgagtc cttttattta aagaagtttg gttctttaaa 540  
 tgaaagaagt ttgggtcttt tgtggatact ctgcatatag accttctaga ctggagtga 599

&lt;210&gt; 120

&lt;211&gt; 198

&lt;212&gt; DNA

&lt;213&gt; Pongo pygmaeus

&lt;400&gt; 120

gaagtttggg tcttttggg atactctgca tatagacctt ctgactgga gtgatatac 60  
 ttgttcagg tcttatgtt ttaaaatgat aactgacac acctaccatt gccaattttt 120  
 aggatatata tctgtgtgaa ctaaggaaag ttcaggaatg attaaacct aatattttgc 180  
 caggagagac tgataat 198

&lt;210&gt; 121

&lt;211&gt; 152

&lt;212&gt; DNA

&lt;213&gt; Pongo pygmaeus

<220>  
 <221> misc\_feature  
 <222> 73  
 <223> n=a, g, c or t

<400> 121  
 tgacttagca gcacagtaat ttggctgaaa tgttgatgag ttcagacaag aagtgatga 60  
 tgacaagaca gcntcacggg gaaattctgc cctgaggaag gaatagtttg aggtggaga 120  
 gatgtgtac ccccgccact ctttgtcaaa tg 152

<210> 122  
 <211> 119  
 <212> DNA  
 <213> Macaca mulatta

<220>  
 <221> misc\_feature  
 <222> 70,81,83,85,88,93  
 <223> n=a, g, c or t

<400> 122  
 tgtcagaaaa caaaaactcc taagagaatt aatagaaatt atcttaagga attataaaga 60  
 acccatctcn cagaaattta ntntnttnac ggnaaaatct catgcaaaga taaccaagt 119

<210> 123  
 <211> 146  
 <212> DNA  
 <213> Macaca mulatta

<400> 123  
 tagtacgtcc acattcattc tagatgtgc ctgaagtcct tagatgcctc tcatagttcc 60  
 ttgccatgag cacttttcca acatggctga tgactgcttc aagcctgtaa ggagatctgg 120  
 tctcagttgg ccaagtcttg gataat 146

<210> 124  
 <211> 74  
 <212> DNA  
 <213> Macaca mulatta

<220>  
 <221> misc\_feature  
 <222> 6,37,48,56  
 <223> n=a, g, c or t

<400> 124  
 ctcagnatat tacctccaac ctacaaattt cctgggnaca cactaagnta ataagnagtt 60  
 aatcacaccg tcct 74

<210> 125  
 <211> 229  
 <212> DNA  
 <213> Macaca mulatta

<400> 125  
 tagtacgtcc acattcattc tagatgtgc ctgaagtcct tagatgcctc tcatagttcc 60  
 ttgccatgag cacttttcca acatggctga tgactgcttc aagcctgtaa ggagatctgg 120  
 tctcagttgg ccaagtcttg gataatgtaa cctcatcacc ggagtcacgt ctcacacct 180  
 ttcccatgtt ccattggta gaagtccag cttctgccca cattcaaga 229

<210> 126  
 <211> 202

&lt;212&gt; DNA

&lt;213&gt; Macaca mulatta

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; 127..128,130..131

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 126

tgaaggtctc cctgctgtgc tgagagaggg gcatgctgga agaagagtgg tcagagaagc	60
aacattgctg gctctaaaga cagaggacag gccatgaacc aaggcatgca ggcggcctcc	120
gaaagcnnng ncaggcaaga gaatggactc tgccctggag cctccagaac gcgtgcagcc	180
ctgcccattc cttgctttta gc	202

&lt;210&gt; 127

&lt;211&gt; 82

&lt;212&gt; DNA

&lt;213&gt; Macaca mulatta

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; 6,8,30,52,72

&lt;223&gt; n=a, g, c or t

&lt;400&gt; 127

aaacantncc aacagaaaca aagctttgtn cttgttttgg gaagcaacct cnaatggtaa	60
gatgggggct gngttctaaa ag	82

&lt;210&gt; 128

&lt;211&gt; 243

&lt;212&gt; DNA

&lt;213&gt; Macaca mulatta

&lt;400&gt; 128

gaacattcac cgtggctgct ctgcgctgcc atacaatgtc agattgcact gaggccgcca	60
tgctgtaaga aaccccaagc tgccctgctg aaaagactgg agagacacag caccatcag	120
cttttccagc cttttgggtc ctcccagcta aattacaggc attgtgaagc agagatgaga	180
agagatgaga cctcctgcgt gccctttcca aatcccggtc cagcagaatc atcctacaca	240
ata	243

&lt;210&gt; 129

&lt;211&gt; 727

&lt;212&gt; DNA

&lt;213&gt; Macaca mulatta

&lt;400&gt; 129

gaacattcac cgtggctgct ctgcgctgcc atacaatgtc agattgcact gaggccgcca	60
tgctgtaaga aaccccaagc tgccctgctg aaaagactgg agagacacag caccatcag	120
cttttccagc cttttgggtc ctcccagcta aattacaggc attgtgaagc agagatgaga	180
agagatgaga cctcctgcgt gccctttcca aatcccggtc cagcagaatc atcctacaca	240
atagtaagtt tttacttgaa gccactaatt tggggagtag ttcttatgca gcaaaagata	300
atcagaacac tgcgccatgg caggagctga taaacctgct cactgttacg cgtgtatcca	360
atgactactc acacacacat gacttgcttc actctacttt ctggctttcc actcagcaaa	420
ttttaagaaa tgagaatgtt attcatgaga aattaactca agatgaatat tgttatcgtt	480
aaactcttcg taaatggtaa aatactacat tattgctagt catttttatt taaagaagtt	540
tggttccttta aatgaaagaa gtttgggtct tttgtggata ctctgcatat agatattcta	600
gacatgaaag aacgttgaag acaagaatg ttgaaggcat atagacatga aggaattttg	660
aagttgggtt cttttgtgga tactctgcat atagacattc tagactggag tgatatgtct	720
tgttcaa	727

&lt;210&gt; 130

<211> 214  
 <212> DNA  
 <213> Macaca mulatta

<400> 130  
 acatgaagga attttgaagt tgggttcttt tgtggatact ctgcatatag acattctaga 60  
 ctggagtgat atgtcttggt caagcttcta tgtttttaa atgataactt gacacaccta 120  
 cccaatgcca atttttagga cataaatctg tgtgaactaa ggaaaattca ggaatgatta 180  
 aacctacata ttttccaag agagactgga taac 214

<210> 131  
 <211> 77  
 <212> DNA  
 <213> Macaca mulatta

<400> 131  
 agagcgtgac ggggaaattc tgccctgagg aaggaatagt ttgaggctgg agagatgtgc 60  
 taccctcgcc actcttt 77

<210> 132  
 <211> 306  
 <212> DNA  
 <213> Macaca mulatta

<400> 132  
 taaaatgact tgttcagggc tccttttggg attttaagt agaaaaactc tatttcataa 60  
 taatgttacc acgctaattc aataacactt catagatggt tagccatctc taaaattaaa 120  
 tttttagtgg gaggagcatg gaaaatacca cctatatattg gatttctggt taacacttca 180  
 aaatccatta ttatcataat ttaaaattag atttaaatac aaacctcttt atagaaaagc 240  
 tgataaatgt gattttatatt gtttgcttgt ttttttgacg ggagggatat tccottgata 300  
 ccactc 306

<210> 133  
 <211> 18  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> sequencing oligonucleotide PrimerPU

<400> 133  
 tgtaaaacga cggccagt 18

<210> 134  
 <211> 18  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> sequencing oligonucleotide PrimerRP

<400> 134  
 caggaaacag ctatgacc 18